March 1993

Format and Content Standards for the Electronic Exchange of Legal Information

Henry H. Perritt Jr.
IIT Chicago-Kent College of Law, hperritt@kentlaw.iit.edu

FORMAT AND CONTENT STANDARDS FOR THE ELECTRONIC EXCHANGE OF LEGAL INFORMATION

Henry H. Perritt, Jr.*

ABSTRACT

This article[1] considers whether format and content standards or conventions[2] for legal information are desirable, and if they are, how they should be developed.[3] Increasing compatibility can reduce barriers to electronic filing with courts and agencies, and to wider use of electronic methods for distributing court, agency, and legislative information to practitioners.

Conventions, including standards, permit communications between persons or entities using different structures for representing information. Frequently, the different representation structures result from different products used to create, read, or manage information. Conventions are aimed at facilitating compatibility. Compatibility refers to the ability of A *to send information*

*Professor of Law, Villanova University School of Law, member of the Virginia, Pennsylvania and District of Columbia bars.

[1]This article is an adaptation of THE ROLE OF FORMAT AND CONTENT STANDARDS IN FACILITATING THE ELECTRONIC EXCHANGE OF LEGAL AND ACCOUNTING INFORMATION, prepared for the NAT'L CTR. FOR AUTOMATED INFO. RES. (NCAIR) with NCAIR support. The author appreciates thoughtful comments on earlier drafts and on the subject of information format standards and content conventions from David R. Johnson of Wilmer, Cutler and Pickering; Marc Lauritsen of Harvard Law School; Suzanne Chambliss Neil of MIT's Media Lab; and Anthony G. Oettinger, Professor and Chairman of Harvard's Program on Information Resources Policy.

[2]To some people, the term "standard" means a centrally adopted convention that is imposed involuntarily on suppliers and users. This report considers voluntary, as well as involuntary, efforts to enhance compatibility. It uses the term "convention" to refer generically to any means of ensuring compatibility. "Standard" is used as a narrower term, signifying an *a priori* specification to which suppliers and users adhere.

[3]The article considers the role of standards for data structures and transmission in facilitating electronic information exchange, with particular attention to Electronic Data Interchange (EDI) and Standard Generalized Markup Language (SGML) as the starting points for useful conventions for legal information. EDI and SGML are nonproprietary standards explained in part V.

to B *in a way that* B *can use it. Compatibility, as used in this article, refers to exchanging data, including text.*[4]

Different standards questions are of interest to different communities. Electronic publishers are interested in consumer preferences for product features and in electronic manuscript preparation and markup. Law firms, courts, and agencies are interested in electronic filing.

This article has three major themes. First, it explains how content compatibility, necessarily involving conceptual or semantic objects, presents much greater challenges than format compatibility, which can be ensured relatively cheaply by simply preserving a human-readable image of the information. Second, it concludes that the pace of technology change makes compatibility easier to achieve by post-hoc conventions than by a priori *agreement on comprehensive standards. Third, it suggests that a collaborative effort to define requirements in terms of information elements can improve compatibility because it makes it easier for suppliers to design standardized information representation into their products, as well as being a necessary prerequisite to the formulation of standards in a strong or formal sense.*

Part I explains why standards are important to lawyers. Part II explains how format and content conventions facilitate information exchange. Part III defines an analytical framework for assessing the different kinds of value that can be added to information, value that can be exchanged only through conventions. Part IV considers the special compatibility needs of the legal profession. Part V reviews existing standards pertinent to legal information, and part VI considers standards development activities. Part VII presents recommendations for the future.

The article concludes that standardizing legal information objects may be appropriate with respect to litigation documents, but is neither necessary nor feasible with respect to the full range of textual documents used by the profession.

I. INTRODUCTION

A. Why Should Anyone Care About Standards or Conventions?

The subject of standards or conventions for electronic information formats is relatively arcane, and it is appropriate relatively early in this article to justify reader attention to the subject.

Consider some things that cannot be done today with desktop computers:

- Lawyers can download the full text of any article or case into their PCs, but not with the page breaks, running headers, boldfaced topical titles,

[4]Text documents in which individual characters are transmitted in machine processable form are the main focus. Anything likely to be integrated with such documents also is considered.

Television and radio broadcasts and oral telephone conversations are not considered because they involve recorded speech and images that are not ordinarily subjected to further content-related processing at the receiving end.

headnote tags, or typeface changes that help them browse the case quickly for pertinent material.

- Lawyers and their publishers are excited about the potential of CD-ROM and other optical media, but every publisher uses a different approach; there is no standard way to use a CD-ROM platter.
- There is no standard way of exchanging files over a network or between application software products, so if *A* and *B* want to work together on the same draft article, spreadsheet, or database, they must use the same software environment.
- Archiving electronic files is a relatively straightforward procedure, but there is no assurance that the necessary hardware and software will be available in ten years to read any of the archived files.
- Graphical images are becoming accepted as a useful tool for lawyers, but there is no generally accepted way to exchange such files so they can be edited, unless the same software is used. The figures in this article, for example, could not be transferred between MacDraft, where they were created, and Microsoft Word for the Mac and Word for Windows, in which the text was written.

In each of these cases, the problem is the lack of compatibility and the absence of a widely accepted convention. The utility of electronic information is, in many respects, limited by the conventions problem. Further growth requires thinking about the problem and possible ways of addressing it.

B. The Special Characteristics of Legal Information

Legal information has some special characteristics that distinguish it from other types of information:

- generally accepted analytical frameworks
- decentralization
- professional customization
- formal communications; and
- lack of embedded technology.

The expression of legal information reflects analytical frameworks accepted throughout the profession. There is little disagreement about the elements of the crime of wire fraud, or the proof necessary to establish lack of informed consent for a medical procedure.[5] Moreover, much legal information is communicated in formal ways, as parts of opinion letters or litigation documents.

Despite the widely accepted analytical frameworks, the authority within the legal profession to define information formats is decentralized. Legal au-

[5]Legal frameworks are dynamic, as the law evolves, and may differ from jurisdiction to jurisdiction because of differences in substantive law.

thority is divided among legislatures, courts, and administrative agencies in the states and federal levels. Within this decentralized framework, governmental authority exists to prescribe information formats for reporting, recordkeeping, and public disclosure. Decentralization discourages standardization.

Lawyers make their living in large part by addressing their clients' unique problems in customized ways. Resistance to electronic content conventions may be intense, based on the argument that lawyers must have the flexibility to frame complaints and other litigation documents according to their own habits and the facts of the particular case, with due regard for strategy and tactics. Obviously, forcing complaints and other litigation documents into a preestablished matrix, such as that represented by an EDI transaction set, inhibits flexibility.

The profession's information processing activities do not represent a large enough market to have caused substantial amounts of technology to be embedded in specialized information formats, although some value-added vendors have developed specialized systems for database development and accounting systems. On the other hand, the absence of substantial embedded technology removes one potential impediment to standardization: it removes one major incentive for vendors to oppose conventions for legal information, because there is less stake in the status quo.

Formality and widely accepted analytical frameworks encourage standardization, representing a counterpoise to decentralization and professional customization. Lawyers are gradually adopting document assembly software that puts together certain kinds of standard legal documents according to preprogammed rules and facts entered into electronic forms for a particular case. As such systems are adopted for litigation practice, the resistance of the bar to predefined information objects for litigation documents will decline. Involvement of courts and agencies in developing document assembly software may reduce the diversity of formats resulting from decentralization.

As this technology spreads, it should become more evident that efficiencies can be realized if a computer, running document assembly software, simply communicates electronically with the court or agency computer rather than having to generate a paper form as an intermediate step. The legal environment will thus resemble, in some sense, the environment existing with major federal agency systems, DOT, ICC, and FMC electronic tariff systems, and the SEC EDGAR system, in which the filer simply sends standardized electronic information objects rather than assembling a paper document.

II. USES OF INFORMATION STANDARDS

To support a strategy for greater standardization of some types of legal information, it is helpful to have a conceptual framework for thinking about electronic information standardization in general.

A. Range of Conventions

Standards or conventions are necessary for the exchange of any kind of information, computerized or not. Human language is a type of standard;[6] so are alphabets, number systems, and musical scoring systems.[7] Information technology requires standards for the exchange of information, just as human communication requires a common language. The Morse Code was a necessity for effective use of the Morse telegraph. The need for technical standards to permit long-distance telephone service was a prime stimulus for the formation of AT & T as the single national telephone company, even before either the ICC or the FCC asserted jurisdiction over telecommunications.[8]

This section develops the following argument: compatibility must exist between both computers and humans. Human processing and computer processing can be substituted for each other in the middle levels of a hierarchy of information exchange. Representing images facilitates human processing. Representing objects facilitates machine processing. Object representation relates closely to content standardization. Image representation relates closely to format standardization. Standardization of object representation is inseparable from the difficult question of knowledge representation. It is easier to "punt" on object representation at some point, and translate the information as an image and give it to humans for further processing. A good example is the preference for facsimile (fax) as a means of exchanging information rather than computer-to-computer exchange of machine processable structures. Nevertheless, there is a gradual shift in emphasis away from purely free-form text to various text database architectures, which substitutes machine processing of information objects for some types of human processing of images. Part 3 develops a framework for considering the types of value added to information at the different levels.

1. *Between Computer Systems*

Compatibility is necessary at multiple levels in information exchange processes. The different levels at which compatibility can be achieved are easier to understand if one thinks about the hardware and software hierarchy that constitutes a computing environment. At the most basic level is the hardware, the processor (CPU), the primary storage or memory, secondary storage like disk drives, input and output devices like keyboards, video displays and printers, and communications devices like modems. All other levels involve

---

[6]English is a *standard.* So is Spanish. Speaking through an interpreter is a *convention.*

[7]*See* Aikin, *Future Imperfect: Is MIDI Making the Right Connections?,* KEYBOARD, May 1990, at 56 (history and future of industrywide protocol for music synthesizers); Westfall, *MIDI Reaches Adolescence,* ELECTRONIC MUSICIAN, May 1988, at 66; Czeiszperger, *Introducing Standard MIDI Files,* ELECTRONIC MUSICIAN, Apr. 1989, at 49.

[8]*See* R. GARNET, THE TELEPHONE ENTERPRISE: THE EVOLUTION OF THE BELL SYSTEM'S HORIZONTAL STRUCTURE 64–65 (1985) (role of technical standards in changing original franchise concept for Bell system); *id.* at 128–31 (Theodore Vail's vision of a unified national system).

software, which are computer programs and the data they operate on. The next hierarchical level above the hardware level is the operating system level. An operating system permits the hardware to make internal information transfer between the video display and the keyboard, and permits application programs to access the hardware. Different operating systems can run on the same hardware, as for example MVS and VM on IBM mainframes or MS-DOS, OS/2, and Unix on PC and PS/2 microcomputers. At the next level is application software, programs like Microsoft Word, WordPerfect, Lotus 1-2-3, or Borland's dBASE IV. This kind of software is intended to perform particular tasks such as word processing or database management. At a still more specialized level are applications for use on one of these application programs. An example would be a particular database implementation written in Borland's proprietary dBASE language, or a set of macros written for Microsoft Word.

Complete compatibility is achievable by using the same hardware, operating system, application program, and application. Typically, this kind of compatibility exists within a single work group. When compatibility is achieved by identical computing environments, no standards are necessary;[9] one simply replicates the environment, however *sui generis.*

Standards or conventions necessarily enter the equation at the next level, which involves data used by the lower levels. Here, there are two kinds of compatibility: compatibility allowing exchange of features useful for human browsing and retrieval, and compatibility permitting machine processability. Human processing is considered more fully in the next section. This section contrasts the two types of compatibility to illustrate how higher levels of compatibility between computers work.

Two concrete examples may help. Deficiencies in the first kind of compatibility are illustrated by downloading a U.S. Court of Appeals case from WESTLAW. The characters making up the text of the opinion arrive intact, as do footnotes and certain title or headline information that is capitalized. There is much information, however, that appears in a printed version of the case that cannot be downloaded through WESTLAW. Italic and boldface styles are not available in the download, and neither is the integration between footnote and text referring to it.[10] These features, useful in browsing a case to retrieve pertinent material, can be transferred only if higher levels of computer-to-computer compatibility exist—if, for example, both the WESTLAW database and the user agree to use WordPerfect native formats, or if they adopt a standard, like SGML, that is capable of representing the typestyle information and footnotes.

The second level of compatibility is illustrated by exchanging an electronic tax return, as might be filed by H & R Block with the IRS under the IRS Electronic

[9]One could argue, however, that use of the same computing environments is a convention.

[10]Integration in this sense means attaching the footnote to the text to which it pertains, so that if the text is moved, the footnote moves with it.

Filing Project. The name of the taxpayer, the adjusted gross income, and the amount of tax due are processed automatically by IRS computers. Those computers are able to "understand" that particular strings of characters and numbers represent these quantities only because both H & R Block and the IRS adhere to a convention that defines the structure of the electronic tax return. It is not enough, if both sender and receiver use the same computing environment; they also must follow the same data structure. When they do this, computer-to-computer compatibility exists at higher levels in the information processing hierarchy, without having to hand off the information to a human for further processing.

Two approaches roughly correspond to the two types of compatibility: image representation and object representation.[11] Image representation is more suitable for ensuring compatibility of the first type, for exchanges intended only for human processing. Only object representation can be used at the second level,[12] for exchanges contemplating further machine processing.[13]

Compatibilities at the hardware level are obtained between different hardware products by standardizing electrical signals and the representation of the lowest-level language components. At the lowest level, there must be compatibility between voltage and current levels used to hook up the originating computer to a communications channel, and the channel to the destination computer. There also must be compatibility in the ways the voltage and current levels signify characters or other lowest-level information elements, such as a bit, which together with other bits makes up codes for alphanumeric characters, or signifies whether a point on a video display is light or dark. Then, depending on the nature of the communications channel,[14] there must be com-

---

[11]The report uses the concept of objects and object representation to include the following features: abstract data types for information, using data structures that model real world entities, encapsulating relevant operations or methods along with the data to which they pertain, and representing particular pieces of information as instances of information classes to which they belong.

Object orientation is now something of a fad among computer programmers, suppliers of programming languages and environments, and those who write about computer programming. There are many definitions of object orientation. *See generally* K. PARSAYE, ET AL., INTELLIGENT DATABASES: OBJECT-ORIENTED, DEDUCTIVE, HYPERMEDIA TECHNOLOGIES 103–05 (1989) [hereinafter INTELLIGENT DATABASES] (explaining basic object orientation concepts). Some commentators assert that the future of intelligent text retrieval depends on a shift to object-oriented concept techniques to reduce the need for content formalization. *Id.* at 363.

[12]The distinction between image representation and object representation used in this article is analogous to a distinction between analog and digital representation. *See* A. Oettinger, The Abundant and Versatile Digital Way 26–30 (1985) (research draft, Harvard University Program on Information Resources Policy, Center for Information Policy Research) (contrasting a drawing of a cat with the word "cat" as examples of analog versus digital representation). Professor Oettinger's paper more generally explains how reduced cost of digital processing has encouraged its wider use, providing advantages in efficiency and reliability.

[13]Machine processing may be necessary even if the ultimate user is a human being. This is the case for example if the human user intends to cut and paste material into a word processor for further editing or formatting.

[14]A switched-circuit channel does not need addressing as a part of the message. A switched-message channel does need addressing compatibility.

patibility in expressing the address of the recipient. This kind of hardware compatibility is ensured through standards developed by international organizations and widely accepted by hardware and software suppliers and their users. The most prominent example is the family of specific standards encompassed by the International Standards Organization (ISO) Open System Interconnection (OSI) model addressed in part 5 of this article.

The ISO OSI standard defines seven layers of interconnection. Above the hardware level, the OSI model deals with things like the size of packets on a communications link, the protocol for resolving contention among different network nodes, and multiplexing or other techniques of aggregating messages to utilize available bandwidth efficiently. Acceptance of the OSI standards has been growing. Layers one through five, covering physical and electrical, data, network, transport, and session compatibility are substantially implemented; higher layers, however, are not.

Addressing standards are in their infancy across network boundaries,[15] although they necessarily are established within any particular network. Agreement on magnetic tape and floppy disk formats has been hard to establish through *a priori* standards.[16] One of the breakthroughs in microcomputer use was the evolution of an MS-DOS convention for floppy disk formats in the mid-1980s, based on proprietary Microsoft and IBM technology. A similar degree of standardization has not evolved yet with respect to optical media, although the "High Sierra" format for CD-ROM media is widely accepted at the present time.

2. *Between People and Organizations*

Beyond hardware and software compatibility lies data compatibility, which has more to do with people than with machines.[17] Once hardware and software compatibility is established, there must be compatibility in the means used to communicate content. The same alphabet must be used, as well as a common language. In addition, a common vocabulary must be used within a particular language. Then, efficient use of information content is facilitated by conventions for organization, so that different subjects or textual passages can be identified, and transitions between ideas can be identified through things like paragraph breaks.

The American Standard for Character Information Interchange (ASCII) specifies how characters, numerals, and basic punctuation marks are repre-

[15]The ISO X.400 standard covers internetwork addressing. It is gradually being implemented by some electronic mail vendors, but its verbose data structure engenders user resistance.

[16]Tapes and disks can be used only on a single type of computer unless standards exist for organization of magnetic bits on the magnetic medium.

[17]The IRS Electronic Filing Project example in the preceding section shows that data compatibility also permits higher levels of machine processability.

sented.[18] Once a standard means of expressing the alphabet has been adopted, natural written language conventions permit compatible expression in text.

Two applications of higher-level conventions, introduced in the preceding section, are important. The first relates to improved human browsability and retrieval. The absence of conventions deprives human electronic information users of typographic browsing and retrieval aids because formatting information—such as bold or italic typefaces, or different type styles and sizes—cannot be stored or transmitted. Thus the absence of conventions inhibits wider use of electronic formats for information retrieval. The second application involves machine processability for sorting, selecting, or taking action.

Higher levels of format compatibility begin with file and message formats. Format conventions should include attributes for indicating whether a particular string of characters constitutes a footnote, attributes for indicating how to resolve an external reference in a compound document, and attributes for indicating whether a particular string of characters should appear in boldface, italics, or a particular font. More ambitiously, they also should deal with document attributes associated with meaning or content: the way a major subheading is expressed, and the hierarchical relationship the subheading has with its parent headings and any subordinate subheadings or text.

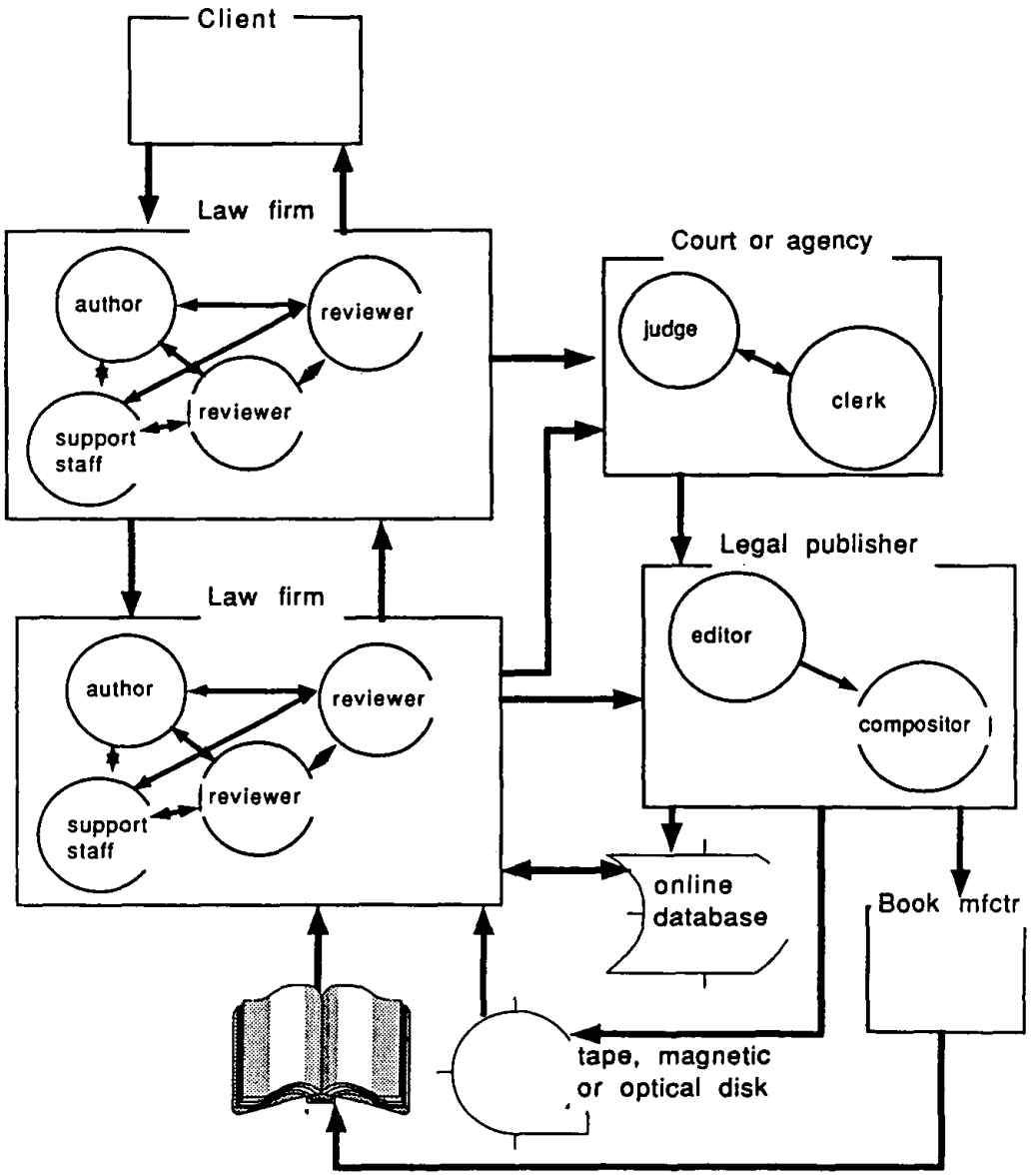
Figure 1 shows the flow of one kind of legal information: litigation documents. Circles show individuals or organizational components involved in processing information (processors). Solid lines with arrows show the flow of information from one processor to another. Thin lines framing rectangles show the boundaries of major organizational entities. Conventions of some kind are necessary for each link between two persons or entities in the information flow. Usually, the compatibility question is neither very visible nor problematic as to information flows within the rectangles, because organizational entities tend to use the same hardware and software and the same conventions for organizing and formatting text. There may, however, be major differences, even within a single organizational entity, as to the conventions used to represent information, and the customs for page layout. Anyone who has exchanged documents with a colleague is familiar with this kind of incompatibility. Compatibility is a bigger problem across organizational boundaries because there is no reason that organization *C* would use the same software and hardware as organization *D*, unless there has been prior agreement, or unless there is only one computing environment.

3. *Object Representation*

Efficient human processing of electronically exchanged information requires format conventions so that typographical features and information structures can be preserved for human retrieval and further human processing.

[18] ASCII is to computer storage and communication what the Morse Code was to telegraphic communication.

Figure 1
Information Exchange Among Legal Entities



Machine processing requires more than format conventions; it requires content conventions.[19]

The boundary between format conventions and content conventions is essentially arbitrary; a convention that relates to format in one context relates to content in another. There is a continuum along which progressively more compatibility is assured by standardizing more and more of the content. For example, making a message machine processable requires standardizing all of the content. An EDI transaction set specifies this kind of standard and a message adhering to a transaction set is completely machine processable. Similarly, a data structure meeting knowledge representation specifications for an expert system or a document assembler has a high level of content standardization. On the other hand, a simple electronic mail convention, simply identifying fields for an address, date, author, subject, and message provides format standardization with little content standardization.

There is inherent overlap between content and format. Essentially, the point of format standardization is to define presentation formats that permit efficient human content-based retrieval. It is useful to think of content standardization as object representation where objects are represented according to preagreed conventions. Exchanging such objects raises no compatibility problems, because the identity and the attributes of each object are expressed in standard and machine processable ways.

Object representation significantly reduces the amount of information that must be stored or transferred to save or communicate an idea. Object representation also permits computer manipulation or editing by the recipient. Consider graphics images. A bit map of a graphical image is neither efficient nor editable. An object representation of the same image is both efficient and editable. One of the simplest forms of object representation in wide use is the ASCII standard for representing characters. Under the ASCII standard, a single character can be represented by eight bits. A bit map representation of a character requires about 45 bits (assuming each character is 5×9 bits, the resolution of the original IBM PC monochrome display). Object representations are machine processable and bit maps are not, but that is only because object representation conventions exist. Virtually every computer system can recognize ASCII characters as such, but generally would not be able to recog-

[19]Professor Anthony G. Oettinger, Chairman of Harvard's Program on Information Resources Policy, suggested thinking about three characteristics of information: substance, process, and format. Substance is content. Different types of substance include entertainment, visual images, fiction, or legal opinions. Process refers to activities used at different stages of handling information, such as gathering, creating, storing, transferring, and converting. Closely associated with the process characteristic are processing components like conduits for information transfer: broadcasting, cable, and mail. Format includes hard copy, electronic visual presentation (CRT), mechanical visual presentation (celluloid film), optical disk representations, and magnetic representations. A. Oettinger, The Information Evolution: Building Blocks and Bursting Bundles 2–3 (Aug. 1989) (Harvard Program on Information Resource Policy No. P-89-5); *see also* ISSUES IN NEW INFORMATION TECHNOLOGY (B. Compaine ed., 1988).

nize bit mapped characters, although there is some trend toward standardizing the representation of character fonts.[20] The choice between bit maps and object representation is presented when one chooses between sending a fax message or an ASCII file, or when one decides between page images or character fields in a large database.

Object-oriented text management approaches are more common than one might think. Every word processor is object-oriented. One defines text to be moved or deleted by "marking" the actual text with the cursor and certain key combinations, not by specifying the starting byte and ending byte and a procedural command. Object orientation is evident when one uses mail-merge codes or style tags and stylesheets in WordPerfect to incorporate variable language. A stylesheet defines certain characteristics for predefined data types, which are objects. Outliners are tools for the creation of flexible object representations of text. The collapsibility feature of an outliner permits one to represent all of the text associated with a particular idea by the entry in a hierarchical position on the outline. Hypertext uses object orientation to link conceptually related textual objects located in different documents. Hypertext requires use of EDI, SGML, or some other text element tagging convention, so the hypertext pointers have useful references.

Higher-level objects can be represented with progressively greater efficiency in the amount of information needed to represent the object. There is, thus, a connection between conventions for information exchange and conventions for "knowledge representation." High-level content standards concern object representation.

EDI transaction sets are definitions of commercial objects like purchase orders and invoices.[21] A specialized document generator used by a particular law office might represent the contents of a sophisticated and lengthy paragraph in a contract by the number 125. SGML[22] concerns itself with representing objects likely to appear in textual documents, beginning historically with the paradigm of a printed page in a book or article.

Standards for legal information, like EDI or SGML, are forms of knowledge representation. Object-oriented representation of a contract requires identifying the attributes necessary to represent a contract. Even more important, in order to make the object machine processable, one must standardize the values that are acceptable in any particular field. Determining whether a particular contract term should be represented by a particular value from a predefined set in, say, the time-is-of-the-essence field, is a fairly high order of decision making about how to represent legal knowledge. Nevertheless, the level of

[20]Standardized fonts equate certain bit patterns with certain objects, for example, a capital P in the Caslon 12-point typeface. Standard fonts thus are a way of moving from bit maps of images to object representations.

[21]The EDI standards are addressed in part V.

[22]The SGML standards are addressed in part V.

decision making about knowledge representation is not beyond reach. The attributes of a contract might be the data elements in an EDI transaction set.

In the future, a significant part of legal information may be created, stored, and retrieved in structures related to meaning. These structures will be created for use in expert systems and other artificial intelligence applications. Any systems for exchanging legal information should take account of these knowledge structures.

4. *Databases Organize Objects*

The difference between a bit mapped image and the encoding of an information object is not unlike the difference between free text and a textual database. Free text is suitable for human content-oriented processing but not machine processing of content. A database is made up of objects in the form of entities, composed of other objects called attributes.[23] Historically, the distinction between databases and free text information was clear. Meaning derived from structure, rather than from context was typical of databases; meaning derived from context was typical of free-form text.[24] A choice regularly confronting designers of document management systems for office automation systems was whether to invest effort at the beginning by defining objects and formats and constructing indexes for eventual retrieval, or simply to save the information in whatever format is best for its original use and burden the ultimate user with the search and retrieval effort.[25]

New digital storage and communication technologies have stimulated a gradual and barely perceptible shift from free-form textual information to text databases. Standardizing formats to facilitate human retrieval and browsing imposes standard structures on free-form textual information. Moving to machine processability requires standardizing content, and that transforms free text into structured objects—the raw material of databases.

The structure of a database is more like the structure of an electronic document—an electronic legal object—than either is like free-form text. The meaning of a particular part of an electronic document, an object, may not be apparent from the context, but only from knowledge of what the contents of a particular field mean. SGML manifests the trend away from purely free-form textual documents toward database representations of textual documents.

Electronic mail is an elementary example because it makes textual information look sort of like a database entry, in that certain addressee, author,

---

[23]Many terminologies are used to describe databases. The entity/attribute model is one. The theoretically rigorous relational model refers to rows, columns, and tables defining relationships.

[24]*See* A. Oettinger, A Bull's Eye View of Management and Engineering Information Systems (1964) (paper delivered at the 19th National Conference of the Association for Computing Machinery) (ACM Pub P-64) (explaining the need for both human-provided context, the "bull," and computer manipulated data, the "cow").

[25]A clear example of this choice is between formatted textual databases and free text search in a law firm context. *See* HENRY H. PERRITT, JR., HOW TO PRACTICE LAW WITH COMPUTERS 596–605 2d ed. 1992.

and subject information is fielded. Growing use of hypertext and hypermedia concepts will increase the convergence between free-form and database paradigms because effective hypertext use requires structuring information into relatively small chunks.[26] The compatibility issues associated with specialized electronic documents[27] structured into objects to permit expert system use are similar to those relating to databases. One can consider the need for standardizing object-oriented electronic documents simply as a part of a larger issue of creating format standards for databases. A "database view" and a "virtual document" really are the same concept.[28]

Compound documents and object-oriented electronic legal documents present somewhat different compatibility problems. The difference between object-oriented electronic documents and databases is that a database usually is composed of many records and is useful in the aggregate. A single electronic document, in contrast, is significantly independent of other electronic documents.[29] Electronic legal documents may lack meaning independent of a look-up table defining the structure of fields (objects) and the meaning of codes in the fields.

Shared databases themselves are ways of communicating between people and organizational components, assuming the requisite standardization between computers. The appropriate standardization of legal objects also can permit the automatic incorporation of an object representing a legal transaction into a database. For example, an electronic tax return becomes a part of databases maintained by the IRS. Document generators or diagnostic systems in law firms or agencies can assemble the files representing particular client matters or transactions and maintain them in databases for later retrieval.

B. Broadcasting and Publishing versus Messaging

Format and content compatibility for legal information are important in two distinct contexts: electronic publishing and electronic messaging, both of which appear in figure 1. Both contexts exist within individual organizations and across organizational boundaries. Electronic publishing involves one-to-many communication, in which the sender of information intends the information to reach many different users. Electronic dissemination of government

[26]Part 3 explains chunking and tagging value and the trend toward changing the size of chunks.

[27]Electronic tax returns filed with the IRS or electronic 10K forms filed with the SEC are electronic documents. The file containing a user's answers to the questions asked by a legal expert system, for example, a will generator or a system for evaluating social security disability claims, also is an electronic document.

[28]C. Dollar, The Impact of Information Technologies on Archival Principles and Practices: Some Considerations 7 (paper delivered at University of Macerata, Macerata, Italy, Sept. 5, 1990).

[29]This may be so, even though the objects making up the document are instances of abstract classes.

information, computer aided legal databases like WESTLAW and LEXIS, electronic treatises, and databases on CD-ROM formats are examples of electronic publishing. Electronic messaging is fundamentally a one-to-one mode of communication in which the sender of information intends to reach one or a few users. Electronic filing, electronic contracting, and the exchange of draft documents among lawyers or between lawyers and clients are examples of electronic messaging.

In the short run, the electronic publishing and electronic messaging contexts present different problems and opportunities for format and content standardization. Electronic messaging may be more susceptible to ad hoc bilateral agreement on information structures, while the one-to-many nature of electronic publishing makes this infeasible. On the other hand, electronic messaging is more likely to involve editing of the received information by the user, and therefore to necessitate information interchange conventions that preserve editability. At least some forms of electronic publishing may not necessitate editable information.

Even if adequate format conventions exist for exchanging electronic *messages,* additional economic, product design, market forecasting, and technical development steps are required to make large scale electronic *publishing* a reality. The availability of CD-ROM and other optical media for distributing large amounts of data—the equivalent of several hundred conventional books—also raises format standardization questions going beyond those associated with electronic messaging.

Even in the short run, however, ensuring compatibility for electronic messaging and electronic publishing presents some similar problems. In both contexts, suppliers (senders) of information would like to reuse existing information in several formats for four different recipients.

Moreover, information structures for legal information are similar in both the electronic publishing and electronic messaging contexts. The structure of an electronically published judicial opinion is not very different from a brief exchanged as an electronic message. The same formats useful for organizing structured information are useful in both contexts. Conventions for these formats and structures can be the same.

SGML is one approach to organizing data on CD-ROM as well as in electronic messages and publishing on electronic networks. SGML also addresses another issue. Electronic publishing includes both the exchange of information on tangible formats like optical or magnetic disks and interchange of information on networks. Economic incentives are strong for publishers to be able to use the same information base for both network and optical disk media.

In the long run, the standardization needs of these two contexts of information exchange merge, because the contexts merge. Electronic publishing on networks is likely to become indistinguishable from, or to merge with, significant aspects of electronic messaging. Billing for electronic publishing requires

structured formats attached in some way to the published material. Hypertext retrieval techniques benefit from smaller chunks of content as retrieval objects. This also results in finer textured structure for published material, the elements of which come to resemble messages. The exchange of information chunks and bills for their use, hypertext pointers, and objects to which they point begin to look like messaging rather than a distinct one-to-many publishing phenomenon.

III. COMPATIBILITY AMONG LEVELS OF INFORMATION

Ultimately, how compatibility is achieved—like the market for electronic information—depends on what consumers want and will pay for.

Some questions relate to electronic technologies used for information that ultimately is presented as an image. A different set of questions relates to electronic technologies that present information to the ultimate consumer in machine processable form. What is the trade-off between human processing and machine processing? What new risks are created by machine processing? What are the costs of new levels of formality necessary to handle the objects necessary for machine processability?

Considering these questions in a principled way requires thinking about discrete value-adding processes in putting together the information product bundle that the consumer sees. Ultimately, of course, technology changes will alter the choices consumers make.[30]

A. Levels of Value in Information

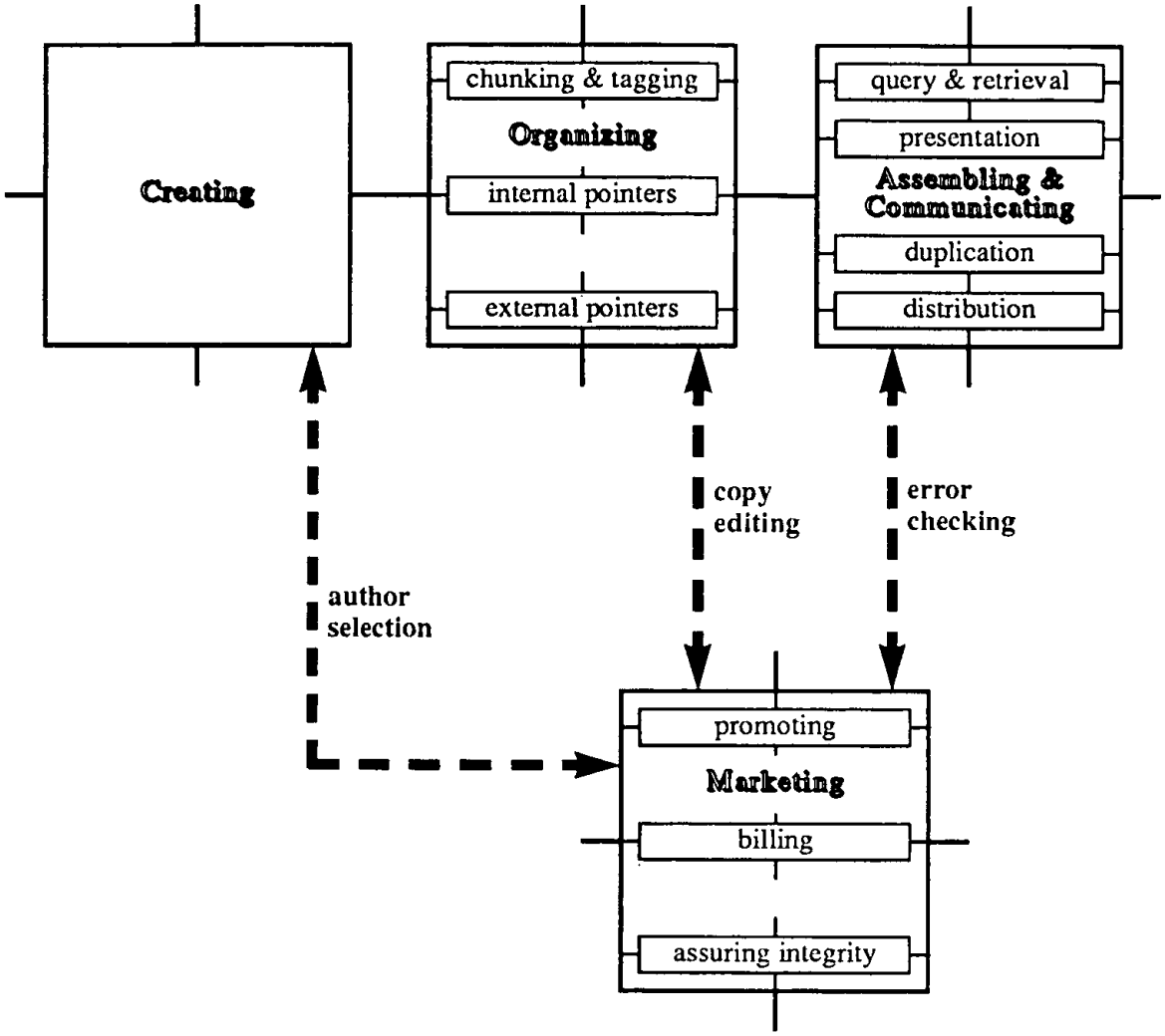
There are four different ways of adding value to information. First, one can create the information. Second, one can organize the information so as to facilitate its later retrieval and use. Third, one can assemble and communicate preexisting information value. Fourth, one can create and communicate information about information. These four ways of adding value represent processes that, like most processes, have inputs and outputs. These processes and resulting levels of value are pertinent to all forms of information—analog graphical images as well as digital electronic invoices for legal services.[31]

An information product, as a user perceives it, is a bundle of intangible attributes and services. The four processes produce 10 different levels of value,

[30]*See generally* Reimer, *2-Pound Palmtops Are Highlighted at Tokyo Exposition,* INFOWORLD, Oct. 29, 1990, at 1 (reporting on Japanese products involving book-sized computers that can accommodate credit-card-sized information cartridges with more than a megabyte of text and graphic information).

[31]The four processes fit text information that is factual in character, as opposed to artistic or functional. *See* U.S. OFFICE OF TECHNOLOGY ASSESSMENT, INTELLECTUAL PROPERTY RIGHTS IN AN AGE OF ELECTRONICS AND INFORMATION 65–66 (1986) (three content categories).

Figure 2
Information Value-Adding Processes and Levels of Value



one or more of which makes up the information product bundle.[32] Figure 2 illustrates the four processes and the 10 levels of value.

The 10 levels represent a kind of à la carte menu from which information product suppliers pick and choose, combining different amounts and varieties of the different levels into products with features the supplier believes will attract significant demand. Electronic information technologies make it easier to separate the different value levels from each other. For example, it is unusual

[32]The 10 levels intersect with the top two layers (presentation and application layers) of the OSI standard.

for an index of a book or its table of contents to be separated from the book itself. It is not unusual for the index or a table of contents menu for an electronic database to be built separately and supplied by an entity different from the entity supplying the underlying data. By thinking about distinct levels of value, it is easier to think about potential compatibility needs and potential opportunities for standardization.

The *creation* process uses as inputs ideas in the mind of the author and information acquired by the author from other sources. It has as output either fixation of information content[33] or oral communication of the content directly to a listener. This process produces *content* value. Content value usually involves the raw ideas fixed in some tangible medium. A rough draft, selection of entities for a database, or a tape containing dictation, has content value but little else.

The *organizing* process uses as an input the output of the creation process. It produces formatted information as an output. The organizing process produces several different kinds of formats. It involves three subprocesses: creating chunking and tagging value, internal pointers value, and external pointers value. The organizing process can be seen as creating receptacles for retrieval hooks, or creating the hooks themselves.[34]

Chunking and tagging value includes typographic features such as: paragraph, section, and page breaks; page numbers; headers and footers; subtitles; headings; tables and summaries; an outline framework; or a database schema.

Internal pointers value includes internal finding-aids like cross-references, tables of contents, indexes, hypermedia pointers, or graphical representations of content.[35]

External pointers value includes external finding-aids like substantive references and citations in footnotes or reference lists, multiple-document indexes, or hypermedia pointers to external objects.[36]

The difference between chunking and tagging value and internal pointers value, both of which deal with internal structure and retrieval aids, is that chunking and tagging is sequential or linear, while internal pointers are random access or nonlinear.[37] Random access capability usually improves retrieval

[33]This concept is borrowed from copyright law.

[34]Tags are receptacles for hooks. Pointers are hooks. Most, but not all, of these features are text *compressors*—short expressions that summarize larger amounts of text to facilitate browsing. *See* R. TAYLOR, VALUE ADDED PROCESSES IN INFORMATION SYSTEMS 59–60 (1986).

[35]Headnotes in reported judicial opinions are a kind of internal pointers.

[36]External pointers value is closely related to abstracting and indexing. *See* TAYLOR, *supra* note 34, at 7 (describing value added by abstracting and indexing operation). Abstracting and indexing creates the hooks by which chunks of information can be linked, through external pointers value to other chunks of information.

[37]Headlines, and other types of prominent topical headings, have characteristics of both chunking and tagging value and internal pointers value. They constitute chunking and tagging value because one must scan the material sequentially to find them. They are random access or nonlinear and thus are internal pointers value because one can read only the headlines without having to read the text between them.

efficiency. The difference between internal pointers value and external pointers value, both of which involve random access, is that internal pointers involve *intra*-document features while external pointers involve *inter*-document features. External pointers include documentary cross-references or bibliographies pointing to other parts of relevant literature, as in legal footnotes citing cases and statutes external to the citing document. Chunking and tagging, internal pointers, and external pointers value all have to do with representation and organization[38] of information to prepare it for retrieval.[39] These can be thought of as *representational factors.*[40] These types of values facilitate retrieval through an assembly and communication process occurring at a later time.

The *assembly and communication* process uses formatted information—the output of the organizing process—as an input and generates information perceived by the consumer as an output.[41]

In highly sophisticated information systems, the assembly and communication process may involve actual machine understanding of an information object followed by appropriate action. This occurs in an electronic contracting system, when a particular value in a particular field of a transaction set triggers some action like the shipment of goods to fill an order or the generation of payment in response to an invoice. In a document assembly system, this assembly and communication process generates the inclusion of a particular language object in response to a particular value in a field.

The assembly and communication process produces several kinds of value.

- *Presentation* value[42] is display of information in a form that can be perceived and used by the intended user.[43] The easiest way to think of

[38]Content, chunking and tagging, and internal and external pointers embrace the "organizing" activity suggested by Kenneth Boulding as one of two knowledge-producing economic activities. The other activity suggested by Boulding is printing. *See* Boulding, *The Economics of Knowledge and the Knowledge of Economics,* LVI AM. ECON. REV. 1, 5 (1966) (capital is knowledge imposed on material world by an organizing process, followed by a process akin to three-dimensional printing).

[39]Attorney David R. Johnson points out that one can distinguish value added to information depending on whether it changes content or whether it affects the way in which, and the ease with which, a reader can find the information.

[40]*See generally,* HENRY H. PERRITT, JR., HOW TO PRACTICE LAW WITH COMPUTERS, ch. 9 (2d ed. 1992) (explaining pervasiveness of knowledge representation issue in automated legal reasoning as an application of Artificial Intelligence techniques).

[41]Consumers of information products may be suppliers to other consumers.

[42]It may be useful to subdivide this level into query and retrieval and presentation, as figure 2 suggests.

[43]The amount displayed may or may not be the same size as the basic "chunk" of information. A database record retrieved may fit on one screen, and a record is the basic chunk in database technology. Conversely, PC word processing software displays about half a page of single-spaced text, and that is smaller than the basic chunk of a textual document—which is the document itself, or perhaps a page, section, or paragraph of the document. *See* TAYLOR, *supra* note 37, at 11 (discussing chunk concept). Attorney David Johnson argues strongly that the paragraph is the natural chunk.

presentation value is as a screen image or a collection of printed pages. Conceptually, however, presentation value also includes information presented to a computer in a form the computer expects and can act on. Thus, presentation value includes the possibility of machine processability. Whether presentation value is aimed at humans or computers reflects whether image or object representation is involved.

- *Duplication* value is a separate copy of the information.
- *Distribution* value is getting the information to where the consumers are.[44]

The fourth process, *marketing,* creates information about information. Marketing uses as an input the outputs of any of the other processes[45] and, rather than transforming the input, creates a parallel flow of information concerning the input. The outputs of the fourth process may be thought of as meta-information,[46] divisible into three kinds of value: promotion, billing, and integrity assurance.

- *Promotion* value is advertising and promoting the resulting electronic information product.
- *Billing and collection* value is accounting, billing, and collecting prices charged for the use of information.
- *Integrity assurance* value assures users and consumers of the integrity of the information product. This level includes the kind of quality control function traditionally performed by publishers,[47] relating to whether the content of a publication is likely to interest a particular consumer, whether it is likely to be accurate, and whether substantial chunking and tagging, internal pointers, and external pointers value

[44]Presentation, duplication, and distribution value all have to do with delivering information, and are in some sense *presentational.* Presentational factors are not independent of representational factors, however. *See* TEXT, CONTEXT, AND HYPERTEXT 36 (E. Barrett ed., 1988) (visual production of text on computer screen is "crippled" without accompanying typographic features); *id.* at 297-99 (screen design for online information products). This raises the question of compatibility between outputs of the organization process and the assembly-and-communication process. The distinction between the organization (representational) and assembly-and-communication (presentational) processes are roughly the same as the distinctions between:

- an information object (representational) and the channel through which it is sent (presentational);
- a product (representational) and supporting services (presentational);
- knowledge production (representational) and knowledge transportation (presentational); and
- processes that make a document what it "is" (representational) and processes that make it available to, accessible to, or usable by someone (presentational).

Marc Lauritsen, of Harvard Law School, suggested these comparisons to the author.

[45]The dotted lines in figure 2 illustrate this use.

[46]Professor Stigler observed that the structure of markets and the role of professions and of other aspects of economic organization frequently are explained best by the need to reduce the cost of information. *See* Stigler, *The Economics of Information,* 66 J. POL. ECON. 213 (1961). Stigler is talking about noninformational products. When the product is information, the kind of information considered by Stigler can be labeled meta-information.

[47]Readers of magazines, journal articles, and newspapers and purchasers of books select material based in part on expectations about the quality of information likely to be published by certain well-known publishers.

has been added, so the consumer can expect high utility. In electronic information formats, integrity assurance value also includes ensuring against corruption and ensuring authenticity.

The shift from mechanical print technologies to digital electronic technologies fundamentally changes the relationship between the organizing and assembly-and-communications processes. The shift makes "publishing on demand" possible, which substitutes chunking and tagging, internal pointers, and external pointers value for presentation, duplication, and distribution value. In publishing on demand, the image is not the medium of search and retrieval.

Figure 3 shows the interaction between organizing and assembly-and-communication in digital publication-on-demand technologies. The organizing processes are represented in the upper left rectangle enclosing three other rectangles. The value-added outputs of those processes are shown to the right of the rectangle, two each for chunked and tagged information, internal index (the same thing as internal pointers), and external pointers. Each of these outputs exists both in on-line form and on a physical magnetic or optical medium, the possession of which can be transferred.

The assembly and communicating processes are shown at the bottom, in the rectangle enclosing four smaller rectangles. There are two outputs, a copy of the assembled information, to the right of the rectangle, and presentation, to the left.

Two links between organizing and assembly/communicating are shown. The one on the left, with its arrow pointing up to the on-line external pointers symbol, represents an on-line query in a network. The one on the right, with its arrow pointing down to the working storage symbol, represents physical transfer of a magnetic disk or tape or an optical disk.

In mechanical print technologies, the drawing would be less complex; the organizing and macro-level assembly would be fixed on a piece of paper. All other assembly and communication would involve that piece of paper.

In electronic publishing, the image is not the medium of distribution, either. Duplication and distribution value is added before presentation value. Less duplication and distribution value is necessary to meet user needs because information content is duplicated and distributed only on request.[48]

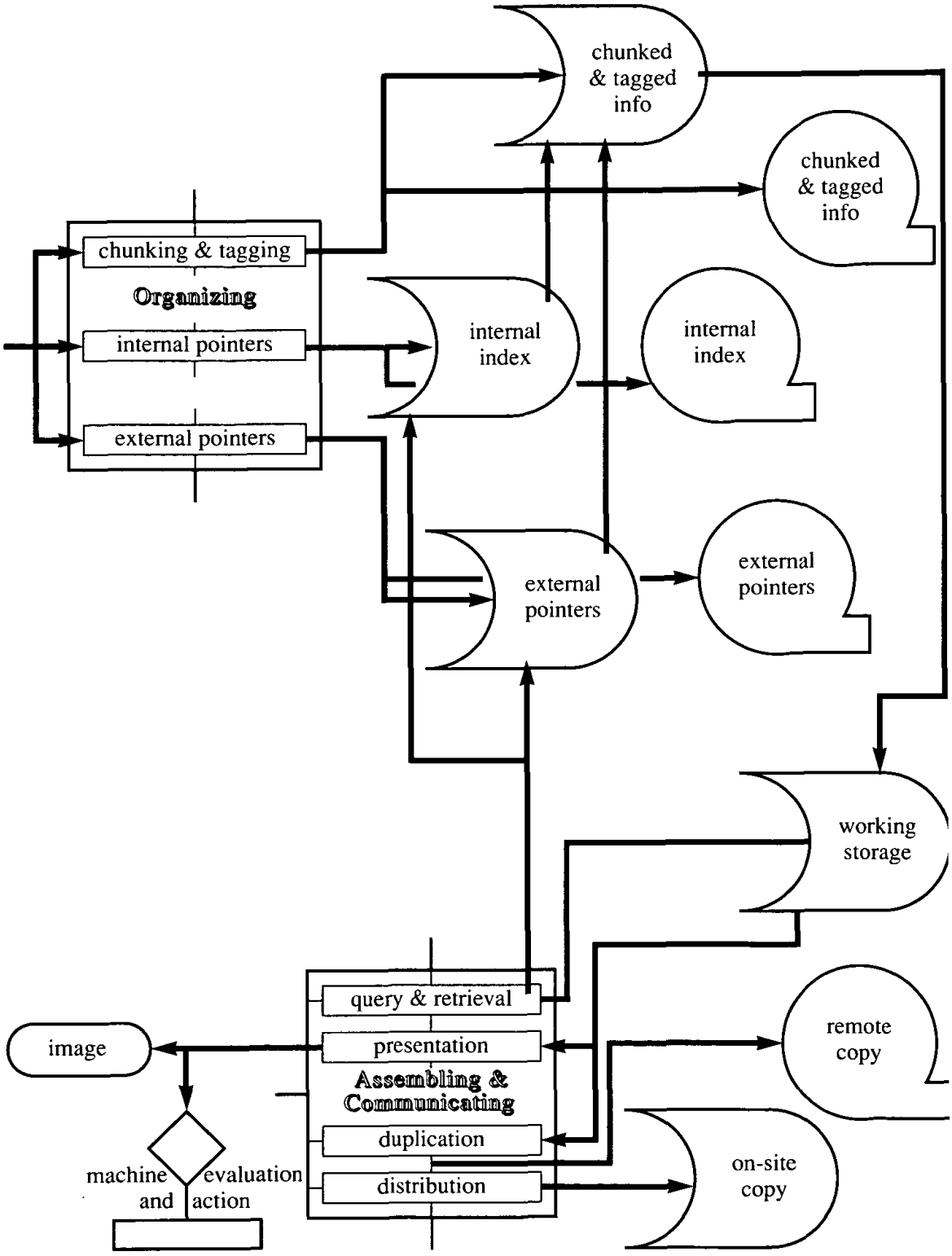
Publishing on demand benefits from smaller units of retrieval. Suppliers substitute chunking and tagging and external pointers value for internal pointers value,[49] necessitating reorganization of content value.[50] Smaller units of re-

[48]Conversely, data broadcasting adds more duplication and distribution value to permit decentralized processing.

[49]When the information object is smaller, the role of the table of contents and index is less, but there are more objects, and thus the role of external tags and pointers to help locate them is greater.

[50]Conceivably an author can write a paragraph so that it would serve equally well to be retrieved as a self-contained unit of expression or to be read in the context of the preceding and following paragraphs. The author of this article is skeptical of the frequency with which this can be done. Most authors must write differently if retrieval of individual paragraphs is contemplated.

Figure 3
Interaction Between Organizing and Assembly-and-Communication Processes



trieval are associated with a less granular representation of information. This is desirable with automated retrieval techniques because a user wants reasonable precision in what he retrieves. There is no point in automating the retrieval process and then having to expend substantial human energy in further browsing and searching within the retrieved object.

Electronic contracting takes automation one step further. Presentation is deferred to the end and involves not a human being, but a computer. When that occurs, the implications for design of all of the levels of value are profound. Object representation is necessary for everything. Human processable images are unimportant.

Perhaps most challenging, as the size and nature of information objects change, the role and nature of the production, distribution, and marketing processes change profoundly. Market structures change as information technologies change. Subsection C considers market structure and organization.

B. Vertical and Horizontal Compatibility

It is important for lawyers to be able to exchange all ten levels of digital information value. Exchanging all the levels of value implicates two kinds of compatibility: horizontal compatibility and vertical compatibility. Horizontal compatibility involves passing information at one value-added level from one person to another, for example, passing a bibliography which contains primarily level-four information. The sender and receiver must agree on the way in which the bibliographic elements are represented. Vertical compatibility involves linking meaningfully associated levels of value with each another. For example, an index (internal pointers value) is meaningless unless it is compatible with the way in which the information object to which it relates is chunked and tagged. The index page number or section number reference must correspond to an actual page or section in the referent object.

Necessarily, as one moves from the lowest to the highest value-added levels, the scope of both horizontal and vertical format standards inquiries changes somewhat. The scope is narrower in the creation and organizing processes and broader in the assembly-and-communication and meta-information processes. The assembly-and-communication process necessitates considering some hardware level issues that need not be considered in the creation or organizing processes.[51]

---

[51] Presentation necessitates considering some hardware level issues that need not be considered at earlier stages. Indeed, the purpose of a page description language standard is to permit interconnection between a variety of user environments on the one side and a variety of presentation devices on the other. The distribution inquiry encompasses lower levels of the OSI hierarchy. One cannot define useful conventions for information interchange in distribution links without addressing multiplexing, packetizing, digitizing, addressing, and error detection and correction.

As to link standards, see Glass, *Moving Toward an All Digital Future,* INFOWORLD, May 14, 1990, at Target Edition 1 (explaining evolution of Bell System T1 Standard—a major standard for high-speed digital transmission). Some commentators believe that two basic options frame

Most of this article is concerned with compatibility for chunking and tagging, internal pointers, and external pointers value. But it is important not to forget the need for compatibility in exchanging the other levels of value.

Even when the image continues to be the medium of retrieval, as in fax communication, conventions are necessary to permit the exchange of typographic features that aid in human search and retrieval: features like underlining, boldface, larger type sizes for headings, pagination, and running headers and footers. Page description languages, discussed in part 4, evolved to permit interconnection between a variety of user environments on the one side and a variety of presentation devices on the other side. It is relatively easy to exchange visual information in the form of a bit map, as long as one knows the resolution of the intended display device, but competing conventions exist for object representations of images. Even with bit maps, conventions are only gradually emerging for compression techniques, to reduce the amount of information necessary to represent the state of an image of a particular size at a particular resolution.

Conventions for exchanging presentation, duplication, and distribution value, resulting from assembly and communication processes are much more mature than chunking and tagging, internal pointers, or external pointers value, resulting from the organizing processes, or promotion, billing, or integrity assurance value, resulting from the marketing processes.

This may seem ironic, because much assembly and communications value depends on compatibility with organizing processes. Nevertheless, standardizing presentation, duplication, and distribution value representations is easier because those levels involve format standardization only, and exchange processes that already are done by computers. Standardizing chunking-and-tagging, internal-pointers and external-pointers value depends on standardizing content that presently is handled by human processing. Standardizing promotion, billing, and integrity assurance value must accommodate realities of market organization, which is itself being changed by the technology to be standardized.

Conventions for exchanging distribution value are highly developed, typified by the OSI standard.[52] OSI is a *suite* of standards, encompassing a number of specific standards, like X.25 for OSI network layer service. There are also emerging standards for exchanging presentation value, in the form of standard bit map, object representations, and page descriptions.

---

the choice of standards for telecommunication links: dedicated T1 (combined with proprietary multiplexers, or ISDN). *Id.* at S4. T1 was designed to carry 24 simultaneous digitized voice channels, with signaling information on separate channels. It operates at 1.544 megabits per second. T1 operates with 193 bit frames. *Id.* at S4.

[52]The OSI standard concerns itself with both vertical and horizontal compatibility. Each layer of OSI is defined so that it can be integrated with the layers below and above it in the OSI hierarchy. In addition, compliance with the OSI standard for any particular layer ensures horizontal compatibility at that layer.

While the OSI concepts are mature compared with conventions for exchanging chunking-and-tagging, internal, and external-pointers value, there are important areas for further development of conventions for presentation, duplication, and distribution. File exchange is a good example. Today, these exchanges usually are done using a ten-year-old file transfer protocol like XModem, YModem, ZModem, or KERMIT. There are better ways to exchange files over wide area network and between wide areas networks. A WordPerfect or dBASE IV file the lawyer transmits to SprintNet or CompuServe must be capable of being transferred from SprintNet or CompuServe to another value-added network that may connect with the addressee. These capabilities are in their commercial infancy, except on the federally-supported Internet and other environments using TCP/IP protocol. The ISO X.400 standard is intended to facilitate this kind of file transfer. Even on Internet, hypertext links and other retrieval enhancements cannot be used on the network.[53]

Standards are lacking for the exchange of information value of each level resulting from the organizing process, except for content. "While the technology to exchange files is well proven, the ability to share information through such files is still quite crude. The only common information representation widely accepted today is ASCII text representation (and even there, such facilities as tab settings sometimes give problems)."[54] Even relatively sequential documents like word processing files require more powerful interchange standards. There is presently no common method for exchanging chunking and tagging value beyond paragraph breaks, and even that is ambiguous because the same code is used to end both lines and paragraphs.[55] There is presently no common method for exchanging internal pointers value in the form of tables of contents, indexes, or outlines. There is presently no common method for exchanging external pointers value in the form of footnotes or hypertext links.

Standardization is not essential with respect to the marketing process because promotion and integrity assurance are qualitative, subjective, and frequently aim to create differentiated products. Standardization would defeat this objective. On the other hand, some types of promotional materials may be more effective if they can be communicated in standard formats or through widely available communications channels. Certain kinds of integrity assurance information relating to authentication, such as cryptographic and sophisticated signature approaches, may necessitate compatibility. And the billing function cannot take advantage of technology unless standards exist to define

[53]In theory, one could set up local translators that resolve pointers into textual retrieval commands in a language that could be understood by a node with the information pointed to, but the network itself has no real capacity to manage distributed data, or to resolve pointers. The "Gopher" convention is emerging as a partial solution on the Internet.

[54]Collaboratory Workshop Report.

[55]*See infra* note 69.

the commercially relevant events (access or use of an information object) and to pass billing information across a complex internetwork.[56]

As the introduction of this section noted, there are vertical dimensions to compatibility as well as horizontal dimensions. It is natural to concentrate on horizontal compatibility. It also is important that each level of information value be accessible by suppliers of other levels of information value. One needs "hooks" for the hypertext links when the creator of hypertext links is not the same person as the supplier of the original data object. A database vendor needs hooks between its user-friendly menus and the data.

The organizing process must be informed by the expected assembly and communications activities. If the assembly and communication process uses screen-size information units as the basic retrieval unit, the information must be organized that way. If information is to be assembled by West Key-number, it must be organized that way.

Any standardization effort must address this kind of compatibility as well as horizontal compatibility, recognizing that the standardization of a format for horizontal compatibility objectives may also achieve vertical compatibility objectives.

Standardization already has attracted some publisher attention in terms of transferring information from authors to publishers, and in terms of transferring information from publishers to ultimate consumers. The focus has also been on reducing production costs.[57]

The AAP SGML effort[58] had its genesis in an effort to enhance vertical compatibility and to reduce costs. If an author submits a manuscript in electronic form and if editors edit the manuscript in electronic form, composition costs may be reduced. More significantly, the author can participate to a greater extent in designing presentation and retrieval formats. A useful standard for author/publisher exchange would accommodate technological improvements in the way authors work. A good authoring tool would work like WordPerfect 5.1, Microsoft Word for Macintosh, Microsoft Word or Windows. The author stylesheets and outline formats would translate directly into hypertext chunks and links, and into graphic design features delivered in electronic or paper products without the need for compositor or human intervention to translate formats, to mark up copy, or to tag or code text for databases.

Other standardization efforts look at the transfer of information from publisher to ultimate consumer. Postscript relates to that type of transfer. But

[56]The billing requirements are less demanding when the information is distributed on CD-ROM rather than across a network.

[57]Conventions for electronic dissemination of court- and agency-originated materials also focus on the authoring function, broadly conceived. The court or the agency is the author. If a publisher can use directly an electronic text generated by a court, its authoring, editing, and composition costs are reduced.

[58]This effort is reviewed in more detail in part 6.

more attention needs to be paid to the influence of consumer preferences on the types of value that authors and publishers add. There is no point in investing major effort in hypertext authoring capability, for example, if consideration of the demand side suggests that lawyers are uninterested in hypertext formats, or, more likely, that they are interested in only certain kinds of hypertext features. The need is for electronic manuscript tools that closely match the final product features desired by lawyers.

Certain electronic messaging applications focus attention more specifically on the demand side. An EDI message is useful only if the recipient can understand and process it, so the development of EDI standards clearly must take into account recipient as well as publisher requirements.

C. Market Organization

Electronic technologies not only change what is available to consumers, they change the costs and allocation of responsibilities among author, editor, compositor, proofreader, manufacturer, advertiser, accounts receivable, and credit activities. Under mechanical print technologies, publishers provide the technology and the organization to connect authors with consumers of their ideas.

Market structures are likely to change as information technologies change. The effect of the shifts from the mechanical print technologies to digital electronic technologies summarized in the preceding paragraphs are likely to fragment the number of suppliers of content and chunking and tagging value, while concentrating the supply of most of the other forms of added value because of the economies of scope of databases and the economies of scale for telecommunication systems. With the shift to electronic technologies, supplier configurations will trend toward a large number of competing vertically-integrated suppliers of content and retrieval-tool value, a smaller number of competing vertically-integrated suppliers of database access, and a still smaller number of competing vertically-integrated suppliers of communication network services.

In such a changed market structure it is far from clear how different suppliers should get paid, how the different works should be promoted, or how integrity assurance should be handled. In other words, it is far from clear who should supply promotion, billing, and integrity assurance value.

Figures 4 and 5 illustrate reorganization of suppliers of value in electronic publishing. Figure 4 is process-oriented, showing multiple suppliers supplying electronic information to a single on-line database. This is the model of WESTLAW, LEXIS, CompuServe, and Prodigy. Figure 5 is product-oriented, and shows how suppliers might be organized along the production and distribution chain. The bull's-eyes represent bundles of added value, according to the value-added framework developed in this part.

Standards and conventions and their use also affect the evolution of market

Figure 4
Multiple Suppliers of Information Value

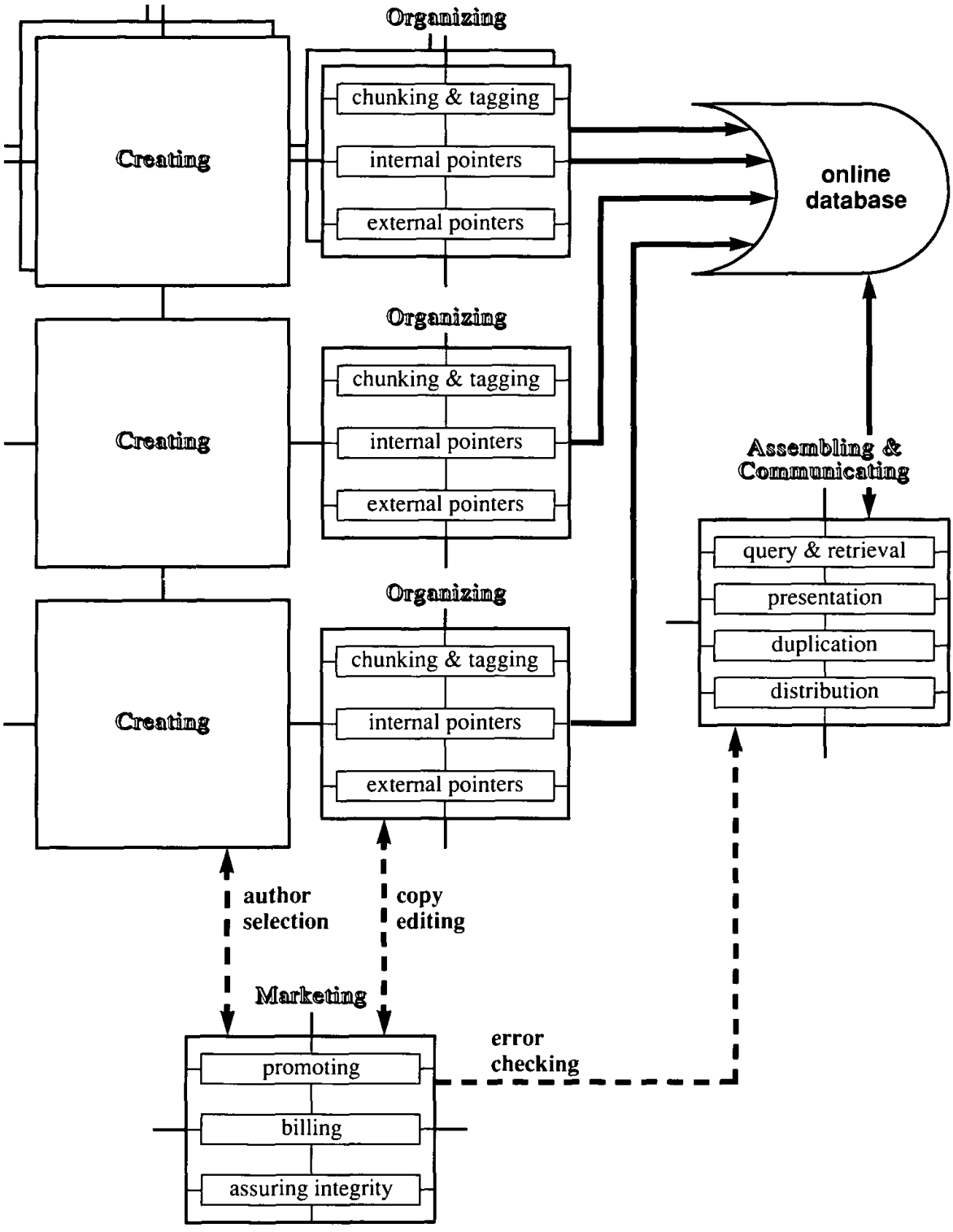
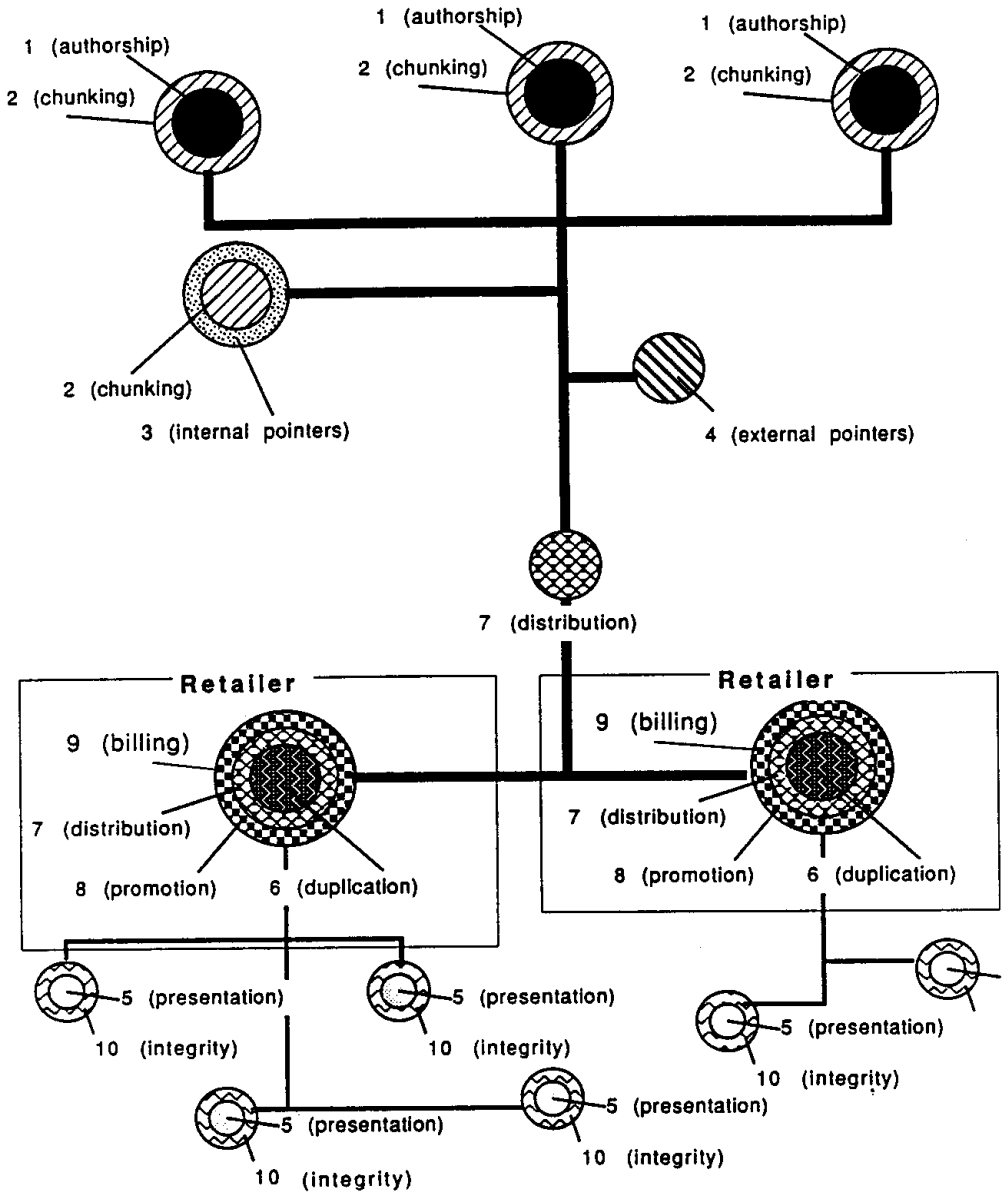


Figure 5
Electronic Publishing—Suppliers



structures. Absence of conventions for information exchange imposes higher costs across firm boundaries,[59] increasing the incentives for vertical integration. More, or more powerful, conventions reduce interfirm transaction costs and thus may lead to less vertical integration.[60]

In the absence of effective conventions for exchanging all ten types of value, vertical and horizontal economies of scope will be substantial because of the transaction costs of exchanging incompatible electronic information formats with other suppliers. Effective conventions for intersupplier transfer and for ultimate delivery to consumers would reduce economies of scope[61] and diminish the incentives for vertical integration, yielding larger numbers of competing suppliers, each seeking market share by innovation in product features. The same phenomenon is likely in terms of standardization's potential to reduce product differentiation. Reduced product differentiation reduces the likelihood of adjacent monopolies and the incentives for vertical integration resulting from that condition.

IV. MEETING THE NEEDS OF THE LEGAL PROFESSION

A. What a Legal Information Convention Should Contain

Because of the close relationship between information exchange standardization and object representation, the first step in developing any standard for the exchange of legal information is to define the information objects to be exchanged. Object definition is a hierarchical activity. For legal information, the object definition process should begin by defining families of documents or object classes.

The two basic families of documents or object classes might be litigation documents and counseling documents. Similar lists of object classes or document type can be developed for the electronic publishing context, working perhaps from basic categories of primary materials and secondary materials.

[59]The costs are likely to increase across organizational boundaries because organizations usually standardize internally.

[60]A profit-maximizing firm is likely to seek vertical integration (producing a factor of production itself rather than purchasing it in the market) to avoid transaction costs resulting from dealing in the market. *See* H. HOVENKAMP, ECONOMICS AND FEDERAL ANTITRUST LAW § 7.2, at 193 (1985). When economies of scale exist for producing the factor, however, the original firm's need for the factor may be insufficient to realize the scale benefits, and it may be cheaper to buy the factor in the market from a supplier who sells to many customers. This occurs when the firm contemplating vertical integration demands less of the factor than the Minimum Optimal Scale (MOS), the smallest production unit capable of achieving relevant economies of scale. *Id.* § 1.4, at 27.

[61]*See* Scherer & Ross at 100–02 (explaining economies of scope); *see generally* Teece, *Economies of Scope and the Scope of the Enterprise,* 1 J. ECON. BEHAV. & ORG. 223 (1980) (enterprise scope determined by transaction costs and realization of economies associated with simultaneous supply of inputs common to processes for producing distinct outputs); Teece, *Towards an Economic Theory of the Multiproduct Firm,* 3 J. ECON. BEHAV. & ORG. 39 (1982) (exploring economies of scope for different inputs).

The following discussion develops electronic messaging objects in greater depth because their paper equivalents are more highly structured and it is thus easier to describe the shift from paper to electronic formats.

Once a particular document family or object class has been identified, one drops down a hierarchical level and defines the objects that make up this higher level object class. Specific document types or objects within the litigation class would be:

- pleadings
- motions
- discovery requests
- pretrial memoranda
- proposed findings of facts
- proposed orders.

Specific document types or objects within the counseling category would be:

- opinion letters
- wills
- settlement agreements
- client memoranda.

The process does not stop here, however. There are still lower level (more specific) objects. For example, one could define the components of a complaint as the attorney representing the complainant, the jurisdiction, the name of the tribunal, the name of the plaintiff or complainant, the name of the defendant or respondent, the docket number, the type of complaint (complaint for injunction, complaint for declaratory relief, or complaint for damages), a series of repeating objects representing factual allegations, one object for each allegation, the relief sought, the signature, and the proof of service, which might contain subobjects for each person served.

Then, one should, for each object thus identified, specify the object type from a limited number of predefined types. For example, the docket number may be numeric, and the jurisdiction may be limited to one of a finite number of instances. The allegation objects may necessitate free-form textual entries, although in some kinds of proceedings, like workers' compensation or fair employment practices, certain of the allegations might be limited to one of a finite number of predefined instances. The civil cover sheets prescribed for federal civil complaints, though they do not constitute pleadings, illustrate the finite instance approach to a concept that is closely related to federal court jurisdictional objects and concepts. In general, if one can conceive of a preprinted form with checkmark boxes for the contents of an object on a complaint, that object is well suited to representation by a predefined instances object type.

The utility of generic document objects may be limited. It may be necessary, for example, to define a different kind of client letter object for different legal specialties and different kinds of client problems. Indeed, some of the

challenges of knowledge representation for legal artificial intelligence are presented in defining objects for legal information exchange. Of course, it is not necessary to define all types of legal information objects in order to develop useful conventions for one or a few types.

After objects are defined, there is another step in developing conventions for the exchange of legal information. One must decide how much information value each object should accommodate. For example, should the standard for the complaint object accommodate stylesheets, or automatic outlining, or counters? Should it accommodate footnotes? Should it accommodate italic and boldfaced styles?

Standardizing outline entry and style tags is particularly desirable. It is extremely useful to be able to represent one or more chunks of text by a single subtitle and associated style tag. The style tag indicates the level in the hierarchy, and the outline entry indicates the subject matter. Most currently available conversion programs are unable to convert either outline tags or style tags or definitions. Being able to standardize or convert this kind of information would be useful not only to preserve the utility embodied in a sophisticated word processing document; it also would be of considerable utility in moving information between word processing document, hypertext, and database structures and in moving information into and out of expert systems and document assembly systems.

This description of standards development reveals obvious difficulties in arriving at the "correct" definition of object classes, objects, and subobjects. Few readers of this section will intuitively accept, with confidence, the author's classification of legal information. The more difficult it is to reach intuitively appealing conclusions about the classification of information objects and their components, the more difficult it will be for parties to agree on a convention. Standards and conventions for information content and format embody classifications of types of information.

It almost certainly does not make sense to invest in developing standards for too many different legal objects; rather, experience should be gained with a limited number of legal objects to develop improved ideas as to the development process and the content of the standards.

Given the certain difficulty in developing a consensus for the organization of legal information, it may be more useful to develop a model of the aspects of legal information, analogous to the value-level model of information value, and define the criteria for exchanging information between layers of the model, or horizontally among users of a particular level of value. This permits players to proceed on pieces of an overall problem, without having to agree on everything. The value-level model described in part 3 may be a useful starting point.

B. How Standards Would Be Used

Use of standards or conventions involves two questions: how the standards or conventions are used to organize information in the standard; and how standardized information is transferred and used.

Once object representations have been developed for some legal information objects, attention must be given to the tools that can be used by lawyers and their support staffs for creating and operating on the information objects defined in the system.

To a very limited extent it may be possible to write procedures that will actually process the objects, much as procedures process electronic purchase orders and invoices in EDI transaction sets. Automatic processing is appropriate for retrieval software, for execution of hypertext pointers, and for using other forms of level-three (internal pointers) and level-four (external pointers) value and their associated chunks and tags (level-two value). Automatic processing is important for this aspect of electronic publishing and for the limited subset of electronic messaging exemplified by EDI, possibly expanded by MIT Sloan School Professor Thomas Malone's information lens and object lens approaches.[62]

In most cases, however, lawyers will use the information in free-form text. They should be able to create and interpret objects using the same methods that are most familiar to them. In other words, if a standard or convention permits a lawyer or her secretary to use word processing software they have already selected, such as WordPerfect or Microsoft Word, they are more likely to use the standard or convention. This makes it attractive to find ways to use outliner features in existing word processing products to create the appropriate objects within any standard or convention that may be developed. In effect, this approach suggests that a standard or convention would be embodied in a series of public domain templates that would appear as particular types of outlines in the native file formats of major word processing programs such as WordPerfect, or Microsoft Word. Such an approach would reduce the transaction costs for defining objects and interpreting their representation, and thus would make use of the standard or convention much more likely.

Transfer and use of standardized information is illustrated by considering how the SGML and the EDI approaches might be used in expressing a standard for civil complaints. EDI would create a transaction set for civil complaints. The transaction set would have an element for each major component. In the U.S. District Court for the Eastern District of Pennsylvania, the elements would be:

- docket number
- name of plaintiff (a potentially repeating element)
- name of defendant (a potentially repeating element)
- title (e.g., "complaint for injunction")
- jurisdictional allegation
- count (a potentially repeating element)
- factual allegation (a potentially repeating element within each count)
- relief sought (a nonrepeating element at the end of each count)

[62]K. Lai, T. Malone, & K. Yu, *Object Lens: A "Spreadsheet" for Cooperative Work*, 6 ACM TRANSACTIONS ON OFFICE INFORMATION SYSTEMS 332 (1988).

- signature of counsel; and
- certificate of service.

Once the transaction set is defined, the structure would be fixed. If the structure were slightly different for another jurisdiction, as it is for the Philadelphia Common Pleas Court which sits in the same geographic area as the Eastern District of Pennsylvania, a different transaction set would be necessary for that jurisdiction.

A standard for the same kind of information in SGML would be expressed somewhat differently. SGML contemplates a hierarchy of progressively more specific definitions, beginning with a possible definition for a class of information objects, say, all civil complaints, that have similar structures and then a more specific document type declaration for the district court and another document type declaration for the common pleas court. Then, at a still lower and more specific level, SGML contemplates a kind of header in an SGML document that specifies the hierarchical structure for that particular document. All of these type declarations use SGML syntax and concepts. That is what it means to say that SGML is a metagrammar. It is a grammar within which one can declare document structures.

V. EXISTING STANDARDS PERTINENT TO LEGAL INFORMATION

A number of specific standards and conventions that have been developed outside the legal community have some potential applicability to legal information.

Some of the most basic and familiar standards pertain to operating systems and networks. MS-DOS is the almost universally adopted operating system for microcomputers of the IBM PC family. MS-DOS includes a standard file representation convention, but does not include any conventions for representing the elements of documents or databases, or for textual representation beyond that provided by the ASCII standard.

UNIX is an operating system developed by Bell Labs that has, for many years, offered the prospect of compatibility among diverse hardware platforms. Widespread use of microcomputers with the 386 and 486 microprocessor has encouraged system developers to adapt UNIX to desktop computing uses. The need for a user-friendly interface for UNIX has been addressed in large measure by the development and release of X/Windows by MIT, which permits diverse hardware platforms to offer windowing and graphical user interfaces in a UNIX environment. Major UNIX vendors are offering their own implementations of X/Windows. The ISO OSI is influencing the design of networks and other digital telecommunications systems. OSI is intended to deal primarily with the interchange of digital information among computer systems of different vendors. It does not specify how files should be structured for interchange-

ability. OSI compliance is required by Government Open Systems Interconnection Protocol (GOSIP), which became effective in August 1990.[63]

The Apple Macintosh, Microsoft Windows, and IBM's OS/2 represent proprietary approaches to the exchange of textual and graphic formats. Features of these graphical user interfaces permit exchanging, between applications, information with a substantial amount of value added.

Database standards are pertinent to the exchange of legal information because of the trend toward merger of free text document and database concepts, reviewed in part 2. If information is stored and communicated via databases, standardization requires standardizing the means of access to the database, though not necessarily standardizing the details of file storage and other hardware and software implementation matters. Structured Query Language (SQL) is a database language standard embraced by the American National Standards Institute (ANSI).[64] SQL is a programming language designed to implement the relational database model faithfully. ISO Standard 8571, File Transfer, Access, and Management (FTAM), provides the capability to exchange data files and to manipulate data files remotely. It specifies an ISO application layer protocol, capable conceptually of accommodating relational and network databases, as well as ODA/ODIF documents.[65] An Information Resource Directory System (IRDS) is an enhanced data dictionary. A data dictionary is a system database that contains definitions and descriptions of data stored in user databases. The IRDS concept permits a uniform approach to a logically standardized, centralized, shareable database for all catalogued information.[66] Each creator of information might maintain a description of information resources in a directory using a standardized IRDS. Then, a central organization would access the descriptions through IRDS service interface.

Document format standards are approaches to defining information structures for the kinds of documents likely to be exchanged between lawyers, between lawyers and official forums, or between lawyers and their clients. Two of the most widely used document format standards are IBM's Document Content Architecture (DCA), and WordPerfect's native format. The DCA convention is incorporated into many proprietary conversion utilities, but DCA is ten years old and lacks the capability of representing certain format features,

[63]*See* 55 Fed. Reg. 27,666 (July 5, 1990), revising FIPS PUB 146, 53 Fed. Reg. 32,270 (Aug. 24, 1988). For a review of GOSIP, see BYTE, June 1990, at 212.

[64]*See generally* 55 Fed. Reg. 16,019 (NIST regulatory agenda referring to FIPS 127, and intention to adopt draft revisions to ISO 9075 for SQL).

[65]*See* PROTOCOLS STANDARDS AND COMMUNICATION, INC., THE APPLICATION OF ODA/ODIF STANDARDS 8 (Mar. 1988) (prepared for the National Archives of Canada) [hereinafter CANADIAN ODA/ODIF REPORT].

[66]*See* McDonald 1989 Archivist Paper at 5; *see generally* 55 Fed. Reg. 11,424 (Mar. 28, 1990) (NIST release on POSIX, discussing data base management standards SQL (FIPS 127), and IRDS (FIPS 156)).

like typefaces, that are becoming increasingly common.[67] WordPerfect is an attractive practical convention because it has such large market penetration. Nevertheless, this approach is proprietary, and there is no assurance that WordPerfect will maintain its market share over the long term. A third proprietary format, gaining popularity, is Microsoft's Rich Text Format (RTF). All Microsoft products can translate to and from this representation language, and a growing number of competing products, like WordPerfect for Windows, can read and write RTF documents.

Several "open" or nonproprietary standards are also useful. The best known, and universally used is ASCII. The American Standard for Character Information Interchange (ASCII) is a widely accepted standard for representing characters, basic punctuation marks, and Arabic numerals. ASCII is universally available as a character representation standard, and accordingly is useful as a way of representing information in higher level formats, such as Postscript[68] and Microsoft's RTF. The ASCII standard does not encompass anything but raw text, however. It has no inherent capacity to express any levels of value higher than level one (content), and a very small part of level two (chunking and tagging). ASCII text cannot include boldfacing, underscoring, or other typeface information. It also has varying implementations with respect to very basic chunking value-like paragraph breaks.[69]

Electronic Data Interchange (EDI) originated as a family of standards for exchanging numerical information as a part of electronic business transactions. Under the standards developed by ANSI, X12, and other EDI groups,[70] a growing proportion of businesses contract with each other electronically by exchanging prescribed data sets[71] between their computers. A purchase order might be issued electronically, or a request for bids and responding bids exchanged electronically.[72]

---

[67]DCA can define some 65,000 different character sets, which could be used to specify typefaces, though not to define them.

[68]Postscript is discussed in the page description format section of this part.

[69]Some implementations use a single line feed character (hex OA) or a single carriage return character (hex OD) and some implementations use a carriage return/line feed pair to indicate the end of a paragraph. Other implementations use either or both of these characters to indicate line endings and two instances of the line ending indicator to indicate a paragraph break.

[70]Electronic Data Interchange (EDI) occurs in large measure under a set of standards developed by ANSI Accredited Standards Committee X12. EDI generically refers to a group of practices, which are not necessarily limited to any particular standards. Indeed, some of the earliest EDI implementations involved data structures developed by large purchasers or suppliers and extended to their vendors and customers by individual contract arrangements.

[71]These prescribed data sets or data structures are called *transaction sets.*

[72]*See generally* National Institute of Standards and Technology, Second Solicitation of Comments on Proposed Federal Information Processing Standard (FIPS) on Electronic Data Interchange (EDI), 55 Fed. Reg. 28,274 (July 10, 1990). This report describes EDI, and lists several applications:

a. Vendor search and selection: price/sales catalogs, bids, proposals, requests for quotations, notices of contract solicitation, debarment data, trading partner profiles;
b. Contract award: notices of award, purchase orders, purchase order acknowledgments, purchase order changes;

An EDI standard might say[73] that the first 7 bytes are the message number, the next 10 bytes the purchase order number, the next 25 bytes a textual description of the product ordered, the next 6 bytes the total price, and the final 10 bytes the signature.[74]

EDI standards development is moving toward accommodating a more diverse universe of document types. EDI may represent a framework for developing a better database standard, related to SQL.

Standard Generalized Markup Language (SGML) is an international standard with growing support in the federal establishment. A generalized markup language defines styles such as "headline1," "headline2," and "bodytext" in a header, in terms of their typeface, point size, and style. Then portions of text to be set as headline1 are marked with tags like "headline1," or "bodytext." The typesetter or printer driver performs the procedural steps to set the marked text according to the defined styles. Procedural knowledge is separated from the text. Content as well as format can be indicated in SGML. Tags can be defined to identify hierarchical components of the logical structure of a document, functionally equivalent to fields in a database description or data elements in an EDI transaction set. SGML is a standard way to specify the tags and procedural definitions. SGML does not identify or specify "standard" document types, document architectures, or text structures.[75] Rather, it is a language or "grammar" for defining document structures and elements in a standard way. SGML standard document-type definitions are conceptually like EDI transaction sets.[76]

The Association of American Publishers has adapted SGML for author manuscripts, electronic composition and for electronic publishing formats like CD-ROM. There has been some movement toward use of the AAP SGML standard, but it is hardly in universal use.

Office Document Architecture/Office Document Interchange Format

---

c. Product data: specifications, manufacturing instructions, reports of test results, safety data;
d. Shipping, forwarding, and receiving: shipping manifests, bills of lading, shipping status reports, receiving reports;
e. Customs: tariff filings, customs declarations;
f. Payment information: invoices, remittance advices, payment status inquiries, payment acknowledgments;
g. Inventory control: stock level reports, resupply requests, warehouse activity reports;
h. Maintenance: service schedules and activity, warranty data;
i. Tax-related data: tax information and filings;
j. Insurance-related data: claims submitted, claims approved.

[73]The example is greatly simplified from a real EDI standard.

[74]EDI transaction sets also may identify information elements by delimiters rather than fixed length field position.

[75]ISO Standard ¶ 1(a) n.a.

[76]ISO Standard at 66 (¶ B.1.2); *id.* at 71 (¶ B.4.2- allowable structures); *id.* at 74 (attributes and chart resembling transaction set illustrations).

(ODA/ODIF)[77] is regarded by some knowledgeable students of standards as a promising standard for electronic office records.[78] ODA/ODIF permits exchange of documents in both revisable and nonrevisable forms. The nonrevisable form permits display of an image, but not further processing.[79] ODA provides two complementary descriptions of a document: a logical structure, composed of hierarchically arranged content elements like sections and paragraphs; and a layout structure, composed of presentation elements like pages and columns.[80] A document profile accompanying an ODA document specifies character sets, styles, and presentation device requirements. It accommodates compound document characteristics. To date, most of the ODA/ODIF development effort has been European and Canadian rather than American.

Abstract Syntax Notation 1 (ASN.1)[81] is an abstract grammar defined in OSI to permit data structures to be defined for interchange at layers five (presentation) and six (application) of OSI. It thus facilitates exchange of documents in the Office Document Architecture (ODA) and Office Document Interchange Format (ODIF) and permits data interchange in the Information Resource Dictionary System (IRDS)[82] and Remote Database Access (RDA) standards. NIST[83] used ASN.1 for the document transfer prototype software developed in an article for the National Archives and Records Administration because ASN.1 is sufficiently robust to be useful for defining arbitrary and complex data types and values.[84] ASN.1 is useful as a building block for database standards, document format standards, and page description standards.

Page description formats are in wider use than document-oriented standards. The best-known page description is Postscript. Postscript, a product of Adobe Corporation, is a prominent example of a standard that permits interchanging information at the presentation level of value. While proprietary, it permits various kinds of textual and graphical information to be prepared for presentation on a variety of devices, without the user needing to be output-device-specific at the time the output is prepared.

---

[77]ISO 8613-1988. ODIF is a data stream used to transmit ODA-structured documents. CANADIAN ODA/ODIF REPORT, *supra* note 68, at 26.

[78]*See id.*, at Annex A (comparing SGML and ODA/ODIF).

[79]*Id.* at 17; NIST Report Appendix C at 12–13.

[80]CANADIAN ODA/ODIF REPORT, *supra* note 68, at 20.

[81]ASN.1 is a component of the OSI application layer.

[82]IRDS is defined in FIPS 156.

[83]The National Institute of Standards and Technology (NIST) is an agency within the U.S. Department of Commerce charged with developing standards for the United States. While NIST historically deferred to private standard-setting organizations for the development of most standards, it has a statutory role in defining standards for federal computing and telecommunication systems. Through the federal government's purchasing power, NIST thus exerts major influence on the content of standards.

[84]M. Law & B. Rosen, Framework and Policy Recommendations for the Exchange and Preservation of Electronic Records § 2.3.3, at 8 (Mar. 1989) (prepared by the National Computer Systems Laboratory, National Institute of Standards and Technology) [hereinafter NIST Report]. NIST also reports that ASN.1 is expected to be incorporated into SQL within a few years.

Standard Page Description Language (SPDL)[85] is a page description language, functionally like Adobe Postscript. SPDL is not revisable, which has some advantages for archival documents or documents where authentication may be important.

Type fonts are an important aspect of presentation value. Much of the human capital developed over 400 years of printing is embodied in hundreds of different typefaces. Obtaining and tracking appropriate intellectual property permissions is a major consideration in achieving page definition compatibility, because a textual page cannot be defined independent of the means of representing the individual characters making up the text.[86]

Compact disk read-only memory (CD-ROM) is an attractive medium and format for electronic publishing because of its large capacity of about 500 megabytes including indexes,[87] and its low copying cost once a master is made. Disadvantages include slow access times compared with small computer hard disks, and relatively high mastering costs. At the present time, CD-ROM vendors have standardized the physical format, derived from audio disks.[88] Microsoft has defined a set of "CD-ROM Extensions," which permit CD-ROM drives to be used with the MS-DOS operating system as though they were magnetic drives. The early tendency for the drives themselves not to be standard is ending.

There is little agreement, however, on indexing and software standards. Typically, a CD-ROM product incorporates a particular vendor's own proprietary access software and indexing approaches. Optimal organization of infor-

---

[85]ISOJTC1/SC18/WG8N561 (third working draft, Feb. 19, 1988).

[86]Intellectual property protection for typefaces is cloudy. Typefaces provide a useful example of the difficulties engendered by the expression/idea distinction in copyright law and the closely related exclusion of useful processes and methods from copyright. *See generally* Atari Games Corp. v. Oman, 888 F.2d 878, 883 (D.C. Cir. 1989) (while typefaces are not copyrightable, their arrangement in a distinctive fashion may qualify; reversing refusal to register video games as audiovisual displays); Readers' Digest Ass'n v. Conservative Digest, Inc., 821 F.2d 800, 806 (D.C. Cir. 1987) (same; design of cover qualified for copyright).

Typefaces are not copyrightable, according to the legislative history of section 102. 17 U.S.C. § 102 note (excerpting from H. Rep. 94-1476, *reprinted in* 1976 U.S.C.C.A.N. 5663) (deferring protection of typefaces); 53 Fed. Reg. 38,110–38,113 (Sept. 29, 1988) (Copyright Office notice that digitized representations of typeface designs are not copyrightable, but original computer programs to control digitization are copyrightable); *see also* Leonard Storch Enters., Inc. v. Mergenthaler Linotype Co., 208 U.S.P.Q. 58 (E.D.N.Y. 1980) (reviewing difficulty in protecting typefaces through patent, copyright or state unfair competition; dismissing claim of infringement of photographic typefaces).

Typefaces are theoretically patentable, but there are large practical difficulties because of similarities of different suppliers' faces. *See* Leonard Storch Enters., Inc. v. Mergenthaler Linotype Co., 208 U.S.P.Q. 58 (E.D.N.Y. 1980) (patent protection has been problematic because differences from prior art are subtle and difficult for a nonexpert to assess); Itek Corp. v. Information Int'l, Inc., 476 F. Supp. 1043 (D. Mass. 1979) (dismissing patent infringement claim; patent related to phototypesetting invalid because claim was merely a combination of preexisting elements).

[87]In comparison, a 3.5-inch diskette has a capacity of 1.5 megabytes.

[88]This is known as the "High Sierra" format, covered by ISO 9660.

mation on a CD-ROM platter depends on the nature of the information and its use. A designer wants to place the most frequently used information to reduce seek times, and to place information close to other information likely to be accessed before or after, for the same reason. Optimum placement of indexes depends on whether disks will be used by themselves or simultaneously with other disks containing related information.[89]

A number of competing proprietary formats are used for legal information on CD-ROM, including West's PREMIS Format, Matthew Bender's format aimed at organizing secondary materials (rather than serials like West's articles and statutory products), and Folio Corporation's Folio Views.

VI. STANDARDS DEVELOPMENT ACTIVITIES

Standards can emerge in one of two ways: adopted by a neutral body, either governmental or private, or out of a natural practice of following the lead of a product supplier commanding substantial market share.[90] The latter approach results in what is usually called a "de facto standard." This part reviews the dynamics of standards-setting, applying a model of those dynamics to the conditions that exist in the market for legal automation products. It concludes by reviewing a number of standards-setting activities that may be useful "homes" for further efforts to develop standards or conventions for legal information.

A. Dynamics of Standards-Setting

Two theoretical models of information technology standards-setting are useful. The first is one developed by Besen and Saloner.[91] The second employs the BATNA concept popularized by Professor Roger Fisher of Harvard Law School. This section develops a synthesis of both concepts.

Commentators Besen and Saloner[92] have identified two variables that

[89] *See generally* Watson, Noreault, & Turtle, *Designing a CD-ROM Information Structure, in* C. SHERMAN, THE CD-ROM HANDBOOK 243, 264–65 (1988).

[90] Recall that standards are only one way to assure compatibility. A variety of other compatibility-enhancing conventions are possible.

[91] *See also* Michael L. Katz & Carl Shapiro, *Network Externalities, Competition, and Compatibility*, 75 AM. ECON. REV. 424 (1985) (developing model for evaluating incentives for software and hardware developers to achieve compatibility in terms of network externalities). Their analysis of network externalities shows that suppliers with a commanding market share usually oppose standardization of anything except their own product features. The consumer advantages from network externalities are available from the use of the dominant product and not from other products. A standard that encompasses competing products would reduce the advantage enjoyed by the dominant producer. Moreover, legal consumers enjoy the benefits of standardization when there is a dominant product, such as WordPerfect.

[92] *See generally* CHANGING THE RULES: TECHNOLOGICAL CHANGE, INTERNATIONAL COMPETITION, AND REGULATION IN COMMUNICATIONS (R. Crandall & K. Flamm eds., 1989) [hereinafter CHANGING THE RULES]; *see also* Besen & Saloner, *The Economics of Telecommunications Standards, in* CHANGING THE RULES 177, *supra*.

influence the standards-setting process and its outcomes. The first variable involves economic incentives for agreement on a standard. Incentives are the difference between the costs and benefits of a standard. Incentives are low when the transaction costs of standards development swamp the benefits, as with weights and measures.[93] Incentives also are low when standardization eliminates a competitive advantage, easing entry by new competitors and easing customer shifts to those competitors.[94] The second factor is the intensity of disagreement, on the merits, about the content of a standard.[95]

Besen and Saloner identify four cases, representing permutations of the values of the two variables. In the first, "pure coordination" case, incentives are high and there is not intense disagreement about the content of the standard. Most standards set by standards-setting organizations involve the pure coordination case.

In the second, "pure public goods" case, incentives are low, and disagreement about content of the standard also is low. Establishment of standards for weights and measures, time, and language involve the pure public goods case.

In the third, "pure private goods" case,[96] incentives are low and disagreement about content of the standard is high. Stalemate in public or private standards-setting efforts frequently involves the pure private goods case. Sometimes, a de facto standards-setter, like Ashton-Tate or Lotus, tries to prevent other firms from using its proprietary technology. When this happens, the supplier with incentives to prevent spread of the standard uses intellectual property concepts to convert what otherwise would be a public good into a private good, excluding others from using the private good.

Ultimately, whether consensus can be achieved in private cooperative standards-setting depends on whether a small number of participants can prevent an effective standard from emerging, and whether side payments are possible.[97] The environment for legal standards efforts produces Besen and Saloner's "conflict" case,[98] where incentives are high, but disagreement about standards content also is high. Conflict can be resolved, of course, frequently by the incentives resulting from the need to be compatible with the market leader.[99]

---

[93]Besen & Saloner, *supra* note 92, at 178.

[94]*Id.* at 179.

[95]*Id.* at 180 (citing disagreement about VHS and Beta videocassette standards).

[96]The name comes from the fact that, in this case, an acceptable standard will favor a particular product. Adoption of MS-DOS or Postscript as de facto standards involved the pure private goods case.

[97]Besen & Saloner, *supra* note 92, at 185.

[98]Incentives are high, but disagreement on the merits also is high.

[99]Beson and Saloner associate with the fourth case the emergence of the IBM Personal Computer and the MS-DOS operating system as the industry standards for microcomputing from 1984--91. When there is no industry leader in terms of market share, conflict resulting from eager participation by all interested parties in the standardization process can be resolved by side payments and the formation of coalitions, such as has occurred with the emergence of two competing drafters of UNIX standards. Besen & Saloner, *supra* note 92, at 183.

The process of developing standards is a negotiation process, and its dynamics are no different from other negotiation processes. The BATNA[100] concept is a useful conceptual principle for understanding the dynamics of the standards negotiation process.[101]

BATNAs and thus incentives in standards-setting are determined to a considerable degree by market structures.[102] A large degree of vertical integration by suppliers of automation products reduces the incentives for standardization.[103] For a niche supplier of a level of value that must be bundled with other levels of value in order to obtain significant market share, adherence to standards is an absolute necessity. Such a supplier will sell little product unless his product can exchange information easily with other products.[104] This necessity almost certainly will override any transaction costs of participating in standards-setting activities. A vertically integrated supplier, on the other hand, faces no such necessity and has less reason to favor standards. The incentives for such a supplier to agree on standards diminish the greater the independence and self-sufficiency of its product.

Thus, consistent with Besen and Saloner's model, a highly fragmented market structure produces large benefits from standardization, although the transaction costs of getting together on a standard may be high. In such a market structure, providing a neutral forum for working out a standard may have a high probability of producing agreement. Alternatively, if the market has one dominant vertically integrated supplier, that supplier's proprietary formats may emerge as a de facto standard.

Conversely, if there are two or only a few vertically integrated competitors, the likelihood of agreement on the standard is small, because no competitor needs to follow its competitors' standards in order to survive, and neither one wants to yield competitive advantage to its competitors.

In automation of legal information, most suppliers of word processing software offer fairly self-contained products. Each of the major producers has significant market share, and the differences in format structures between some

---

[100]Best Alternative to Negotiated Agreement.

[101]BATNA says that no disputant will agree to an outcome that is worse for that disputant than the best alternative the disputant could obtain in the absence of agreement. If I want to file a lawsuit, and am negotiating with you over a contingent fee, I will not agree on 40% if I can get identical legal services from someone else for 33%. Anyone's position at any point in time in any negotiation thus is determined by that person's perception of what she can get if the negotiation fails.

[102]Besen & Saloner, *supra* note 92, at 197 ("excess inertia can arise if the first potential adopters to consider the new technology are not prepared to give up the transient benefits from being compatible with the installed base of the old technology. They then adopt the old technology, swelling the ranks of the installed base and making the old technology even more attractive").

[103]In the absence of an agreed-upon standard, a vertically integrated supplier can sell self-contained multifunction products into a relatively broad market. Such a supplier's BATNA on a multivendor standard is not unfavorable.

[104]Inability to work in conjunction with other products will produce buyer resistance and reduce the potential market. This supplier's BATNA is unfavorable.

of them, especially between WordPerfect and Microsoft, are substantial.[105] Microsoft Word has an architecture in which text enhancements are separated physically from the text. WordPerfect embeds format codes in the text, adjacent to the affected textual passage. This difference in architectures leads to sharp disagreements about the most appropriate technology, reflects earlier disagreement, and affects incentives.

Most suppliers of vertically integrated word processing products do not, however, have significant market share for database or telecommunications products. To the extent that new standards primarily address exchanging information across boundaries among word processing, database, and telecommunications products, the incentives for existing vendors to adopt standards may not be very great, because their BATNAs are favorable and the transaction and the transaction costs of agreeing on a standard also are high.

The importance of compatibility across the boundaries may increase, however. The rising importance of telecommunications and electronic publishing increases incentives to improve compatibility across the traditional product boundaries. These changes in market demand increase the importance of compatibility between levels of value beyond level internal pointers. They increase the importance of compatibility between level-four (external pointers) and lower-numbered levels of value. They increase the importance of compatibility between level-five (presentation) value and lower-numbered levels so that screen images and printer output can express the full range of the lower-numbered levels of value, and they increase the desirability of compatibility between level-seven (distribution) value and lower-numbered levels so that network publishing can express the full range of value added at lower-numbered levels.

Thus, the relevant market[106] for standards involves the major suppliers of word processing products, the major suppliers of telecommunication products, and the major suppliers of database products. It is unlikely that the incentives to agree on standards spanning the scope of all types of added value would be favorable, and the potential for disagreement about technological merit is enormous. Whether alliances form among these suppliers is a critical question. Whether consumers articulate—verbally, or through their purchasing behavior—the desire for compatibility across traditional product boundaries is an even more critical question.

For example, as networked computers become more common, WordPerfect may see interchangeability of file formats as helpful in retaining or enlarging its market share. WordPerfect generally has disclosed its file formats to help add-on product suppliers interface with the WordPerfect file format, so there is every reason to expect that WordPerfect would cooperate with other suppliers

[105]The BATNA for each is favorable because each has substantial market share and uses a technology it believes is superior.

[106]"Market" in this context simply means the universe of affected suppliers.

seeking compatibility with the WordPerfect format. On the other hand, WordPerfect would have very little incentive to agree on file formats other than its own, because of the adverse effect on its large share of the legal markets.

B. Achieving Compatibility Without Standards

Standards are not the only way to achieve compatibility, if standards are understood to be *a priori* prescriptions for document and database content structures, file formats, and communication streams that must be adhered to by commercial product designers. Conversion is another alternative. Many software products permit users to save information in a number of different formats. Indeed, this feature exists in WordPerfect 5.1, for earlier WordPerfect formats, for IBM's DCA/RFT, and for "text" (WordPerfect term for ASCII). Each option involves converting WordPerfect's proprietary formats for version 5.1 to the selected output format. Similar options exist in Microsoft Word. Borland's Quattro permits saving spreadsheets in other vendors' formats. There are many other examples.

Conversion software also exists separately from application products. There are major limitations with both of these approaches. Currently, for example, none of the conversion products with which the author is familiar translates style sheets. Many of the conversion products and the file format saving features eliminate footnotes in the conversion process—a major disaster for law review articles, and a minor disaster for briefs. The current state of the art is incapable of facilitating the transfer of much more than rudimentary chunking and tagging value.[107] They do, however, permit more value than can be represented in the ASCII standard to be translated.

Moreover, frequent conversion of files historically has been inconvenient. It may be acceptable to convert a document once or twice, say, between the final draft on one system and the typesetter's system. It is less acceptable to convert with every use, as between a lawyer's desktop computer and a supporting secretary's computer, unless the conversion programs are very good.

On the other hand, the record of vendors of conversion software in keeping pace with technological innovation is much better than the record of standards developers. Typically, conversion products lag by about six months behind the introduction of new versions of popular products.

Standards development lags much longer. Developing *a priori* standards with any prospect of success requires suppliers and consumers to agree in advance on content hierarchies, available formats, and implementation details. Such agreement is possible only with much effort, and only in limited circumstances.[108] Consider the prospects for developing a general standard for legal

[107]Stylesheets are a kind of level-two value which is not translated.
[108]Part VI, Section A explains the dynamics of standards-setting.

information, one that could encompass the full range of legal filings, from a summary judgment memorandum in a personal injury lawsuit, to a financing agreement for a Boeing 767, and also handle electronic commercial documents like a funds transfer order between large banks or an invoice for medical supplies. The intellectual task of designing a sufficiently general information structure is daunting, and the politics of such an effort would be insufferable.

There are, however, other ways to promote useful levels of compatibility. One is to rely on translators. Another is to develop *a priori* standards only for limited domains, like filings for specific forums, or specific types of two-party relationships, where an assessment of the dynamics under the framework offered in part 5 is promising. These approaches pertain more to messaging than to publishing. In publishing, it may be feasible to develop standards for certain levels of information value, with exchanges at other levels handled through more flexible conventions, such as copying files from optical to magnetic media before routine use. Other conventions might include standardizing hardware and software environments, but this has adverse implications on market size and on utility.

C. Specific Activities

The preceding section explained that the dynamics of standards-setting make emergence of a multivendor standard likely under only certain market structure and technology circumstances. Even when the configuration is favorable, some kind of neutral forum frequently is necessary. The following sections introduce some existing activities that might become vehicles for such multivendor agreement.

1. *Judicial EDI*

An informal group to develop judicial EDI standards for electronic litigation documents and dockets has been functioning for more than two years. An initial, informal, public meeting to discuss standards for electronic court filings was held on June 5, 1990, in Alexandria, Virginia, in conjunction with a regular meeting of the American National Standards Institute (ANSI) X12 Committee.[109] There have been several meetings of the group, which is composed of representatives of vendors, the National Center for State Courts, the Administrative Office of United States Courts, and the ABA.

2. *SEC EDGAR*

The SEC EDGAR dissemination contractor is using SGML document-type descriptions for textual information in electronic SEC filings. The ED-

[109]The X12 Committee has developed some 100 Electronic Data Interchange (EDI) standards for commercial documents like purchase orders and invoices, which are in wide use in the transportation, automotive, retail clothing, and health care industries.

GAR effort is the most mature SGML-related format standards effort directly pertinent to information of a legal character. It focuses on electronic publishing rather than electronic messaging. It is, however, focused on the particulars of SEC filings, and whether its results can be extended usefully to legal information more generally is questionable.

3. *Text Encoding Initiative*

The Text Encoding Initiative (TEI) by the Association for Computational Linguistics, the Association for Computers and the Humanities, and the Association for Literary and Linguistic Computing began in November 1987, and has proceeded to the point of draft "guidelines for the encoding and interchange of machine-readable texts," version one of which was released on August 2, 1990.

> This interchange format is intended to specify *how* texts should be encoded or marked-up so that they can be shared by different research projects for different purposes. The use of specific types of delimiters to distinguish mark-up from text, the use of specific delimiter characters, and the use specific tags to express specific information, are all specified by this interchange format, which is closely based on the international standard ISO 8879, *Standard Generalized Mark-up Language* ("SGML") . . . .

The TEI structure includes a header, which identifies the source of a document, those responsible for production, and specifies the details of coding conventions. Tags are specified for prose text components, including parts, chapters, sections, along with headings and conclusions, title pages, forewords, acknowledgments, abstracts, dedications, frontispieces and appendices, glossaries, and notes. The structure also encompasses external references, cross-references, and hypertext links (non-textual components like tables and figures). Version one does not include formal SGML DTDs for all of the tags defined in the TEI structure.

The TEI initiative, like the AAP initiative discussed in this part, illustrates comprehensive SGML-based standards without focusing on the particulars of legal.

4. *Association of American Publishers Initiative*

The Association of American Publishers developed document-type descriptions under SGML. The AAP initiative includes document-type descriptions for generic documents such as books and magazine articles. The appendix includes an AAP SGML document-type description.

5. *Government Roles in Standards-Setting*

The U.S. Government historically has been reluctant to intervene frequently or heavily with respect to technical standards for information processing. Commonly, the government defers to private organizations, chiefly ANSI and IEEE, to represent the United States in international standards-setting activities.

The National Institute of Standards and Technologies (formerly National

Bureau of Standards) has statutory authority to develop standards for the government itself. Increasingly, in recent years, NIST has been authorized or directed to play a role in setting standards for federal agency automation systems. In most cases, exemplified by a number of Federal Information Processing Standards (FIPS),[110] NIST has simply adopted private standards endorsed by ANSI.

There are exceptions to this abstentionist approach, and the record is decidedly mixed as to the success of more affirmative action by the federal government. In the late 1950s, the Defense Establishment took the initiative in developing a high-level language for processing administrative and commercial information. The result was COBOL, which was as widely adopted as any other computer language. The U.S. Navy took the initiative in developing a standard format for textual documents. The result, Navy DIF, has not been followed very widely.[111] More recently, the Defense Department has adopted SGML as a standard format approach for its computerized logistics system. It is too early to tell what impact this decision will have.

Absent additional legislative authority, there are only four ways in which the federal government can promote the adoption of standards. First, the government can use its market power as a purchaser, forcing vendors who wish to sell to the government to adhere to selected standards. Requiring agencies to conform to the standards is the major way in which market power is mobilized on a broad basis. Second, the government can endorse standards without imposing them even on its own agencies. Third, the government can participate as a regular member in standards-setting organizations.[112] This participation presumably lends some additional acceptability to resulting standards. Fourth, a single government agency can require the adoption of certain standards for information filed electronically with it. All of the agencies involved in electronic filing programs[113] necessarily have used this approach—though not the same standard—in connection with their electronic filing initiatives. Imposition of format standards for electronic filings can be preceded by extensive consultation, resembling standards development for voluntary acceptance by vendors, or it can be essentially a unilateral imposition. The politics of electronic filing programs militates in favor of some degree of consultation.

---

[110]Several FIPS are considered in part V in conjunction with an assessment of particular standards candidates.

[111]*See generally* NIST Report, Appendix C at 11, describing National Bureau of Standards, *Document Interchange Format (DIF)*, NBSIR 84-2836).

[112]The third approach is not inconsistent with the first or second; the government could participate in developing a standard, endorse it and recommend its use by the private sector, and require its use by federal contractors.

[113]*See generally* Henry H. Perritt, Jr., *Federal Electronic Information Policy*, 63 TEMPLE L. REV. 201 (1990); Henry H. Perritt, Jr., Electronic Acquisition and Release of Federal Agency Information (Oct. 1988), report supporting ACUS Recommendation 88-10, 54 Fed. Reg. 5207 (Feb. 2, 1989), codified at 1 C.F.R. § 305.88-10 (1990); Henry H. Perritt, Jr., *Electronic Acquisition and Release of Federal Agency Information: An Analysis of ACUS Recommendations*, 41 ADMIN. L. REV. 253 (1989).

Even when the government is not imposing standards for information to be filed with it, governmental involvement makes a difference in two major ways. It adds legitimacy to the standards-setting process and to the standard that emerges. It creates an institutional setting in which there may be personal incentives for the standards-setting process to succeed. It is not uncommon for a senior governmental official to set as a personal goal the development of a particular standard.

6. *Desirability of Further Standardization Activities*

The legal profession should not seek at this time to standardize textual documents in general. The nature and uses of such documents are too diverse, and there is no perceived need for machine readable representations for such a broad standardization effort to get anywhere. It is not particularly desirable, nor is it feasible, to develop a new standard for expressing the basic components of text or compound text and graphics documents. The legal profession lacks sufficient market power to force adherence to such a standard by software and hardware suppliers, and there are no incentives to develop a standard different from the nonproprietary ASCII standard and the widely used proprietary WordPerfect and Microsoft RTF conventions.

The legal profession should, in any standardization effort, work within the SGML framework. SGML is a widely accepted grammar for expressing text and compound document features in a nonproprietary way. Translators exist for converting Microsoft Word and WordPerfect style tags and formatting codes to and from SGML tags.

The legal profession should begin to develop standard ways of representing structured legal documents, beginning with litigation documents. As court and agency automation proceeds, the need for machine readability will increase. While it may not be feasible or desirable for all elements of a litigation document to be machine processable, there are major advantages for making captions, parties, counsel name and address, docket numbers, and document type machine processable. As court automation proceeds, some courts and agencies undoubtedly will develop ad hoc standards for electronic representation of litigation documents. Before this proceeds so far as to make broader standardization difficult, some kind of profession-wide effort is desirable at least to find a limited number of approaches to electronic document standardization. The judicial EDI effort may be an appropriate nucleus, especially if its ties to appropriate parts of the ABA are strengthened.[114]

---

[114]The choice between EDI and SGML as the starting points is important. It represents a choice between static templates and abstract standards languages.

As the section on SGML explains, SGML is a more abstract and therefore extensible approach to textual standards than EDI, which is more static and specific to a particular application—a kind of template. In general, one can ensure growth potential and broader applicability of a standard by making it more abstract. ASCII is a good example. But adaptability comes at a price. More work must be done with an abstract standard to represent real world information than with a more specific standard. If the "standard" is the English alphabet or the Morse Code, the range of

The National Center for Automated Information Research (NCAIR) hosted a major conference on electronic publishing standards for legal and accounting information in May 1991. But an NCAIR working group concluded that efforts to formulate an actual standard by NCAIR was not the best use of NCAIR resources. Rather, the working group, chaired by the author of this article, concluded that NCAIR would be better advised to sponsor a series of demonstration projects encouraging practitioners, vendors (including legal publishers), and legal academics to use more advanced forms of information technology.

VII. CONCLUSION

Content compatibility, necessarily involving conceptual or semantic objects, presents much greater challenges than format compatibility, which can be ensured relatively cheaply by simply preserving a human-readable image of the information. The pace of technology change makes compatibility easier to achieve by post hoc conventions than by *a priori* agreement on comprehensive standards. A collaborative effort to define requirements in terms of information elements can improve compatibility because it makes it easier for suppliers to design standardized information representation into their products, as well as being a necessary prerequisite to the formulation of standards in a strong or formal sense. Standardizing legal information objects may be appropriate with respect to litigation documents, but is neither necessary nor feasible with respect to the full range of textual documents used by the profession.

---

application is enormous, but only the lowest levels of information value have been standardized. At the extreme end of the abstract/specificity spectrum, one could place general purpose programming languages like C or Basic. Simply offering one of these languages as a standard would not advance compatibility of information format or content for legal information much.

Conversely, with an EDI transaction set, all one need do is plug in data. But an EDI transaction set has no flexibility to accommodate changes or local variations in information content or structure. A new transaction set must be developed for each application.

In selecting an approach or a particular candidate for standard, one must deal with the inherent tension between these two approaches, recognizing that there are advantages and disadvantages of each approach.